PATENT
81940.0071
Express Mail Label No. EV 324 112 150 US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	Art Unit: Not assigned
Masahiro ARAI et al.	Examiner: Not assigned
Serial No: Not assigned	
Filed: February 20, 2004	
For: Magnetic Disk Array Device With Processing Offload Function Module	

TRANSMITTAL OF PRIORITY DOCUMENT

Mail Stop PATENT APPLICATION
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Enclosed herewith is a certified copy of Japanese patent application No. 2003-393912 which was filed November 25, 2003, from which priority is claimed under 35 U.S.C. § 119 and Rule 55.

Acknowledgment of the priority document(s) is respectfully requested to ensure that the subject information appears on the printed patent.

Respectfully submitted,

HOGAN & HARTSON L.L.P.

Date: February 20, 2004

By: _____

Anthony J. Orler
Registration No. 41,232
Attorney for Applicant(s)

500 South Grand Avenue, Suite 1900
Los Angeles, California 90071
Telephone: 213-337-6700
Facsimile: 213-337-6701

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: ２００３年１１月２５日

出 願 番 号
Application Number: 特願２００３－３９３９１２

[ST. 10/C]: [ＪＰ２００３－３９３９１２]

出 願 人
Applicant(s): 株式会社日立製作所

２００３年１２月２５日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号 出証特２００３－３１０７４３８

【書類名】　特許願
【整理番号】　KN1577
【提出日】　平成15年11月25日
【あて先】　特許庁長官殿
【国際特許分類】　G06F 12/00
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
　【氏名】　新井　政弘
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
　【氏名】　松並　直人
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
　【氏名】　八木沢　育哉
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
　【氏名】　萬年　暁弘
【発明者】
　【住所又は居所】　神奈川県小田原市中里３２２番地２号　株式会社日立製作所　ＲＡＩＤシステム事業部内
　【氏名】　佐藤　雅彦
【特許出願人】
　【識別番号】　000005108
　【氏名又は名称】　株式会社　日立製作所
【代理人】
　【識別番号】　100093492
　【弁理士】
　【氏名又は名称】　鈴木　　市郎
　【電話番号】　03-3591-8550
【選任した代理人】
　【識別番号】　100078134
　【弁理士】
　【氏名又は名称】　武　顕次郎
　【電話番号】　03-3591-8550
【手数料の表示】
　【予納台帳番号】　113584
　【納付金額】　21,000円
【提出物件の目録】
　【物件名】　特許請求の範囲　1
　【物件名】　明細書　1
　【物件名】　図面　1
　【物件名】　要約書　1

【書類名】特許請求の範囲
【請求項１】
ＡＴＡ磁気ディスクと、前記ＡＴＡ磁気ディスクを制御するディスクアレイコントローラと、前記ディスクアレイコントローラと前記ＡＴＡ磁気ディスクとのパス上に存在するインターフェースカード内の処理オフロード機能モジュールと、を備えた磁気ディスクアレイ装置であって、
前記ディスクコントローラは、リード又はライトを一例とする標準処理を行う標準処理ＦＣコマンドと、ベンダユニークなオフロード処理を行うオフロード処理ＦＣコマンドと、を前記インターフェースカードに出力し、
前記インターフェースカード内の前記処理オフロード機能モジュールは、前記標準処理ＦＣコマンドに対しコマンドマッピングテーブルを用いて対応するＡＴＡコマンドを前記ＡＴＡ磁気ディスクに発行し、前記オフロード処理ＦＣコマンドに対しＡＴＡプロトコルで最適処理となるＡＴＡコマンド群を用意し必要なときには演算する
ことを特徴とする磁気ディスクアレイ装置。
【請求項２】
請求項１に記載の磁気ディスクアレイ装置において、
前記処理オフロード機能モジュールは、コマンド解析処理部と演算処理部を備え、
前記コマンド解析処理部において、前記ディスクアレイコントローラからの標準処理ＦＣコマンドがＡＴＡコマンドにマッピング可能であるか否かを判断し、
前記演算処理部において、前記オフロード処理ＦＣコマンドを前記コマンド解析部から引き継いで前記ＡＴＡコマンド群を用意する
ことを特徴とする磁気ディスクアレイ装置。
【請求項３】
請求項１又は２に記載の磁気ディスクアレイ装置において、
前記ディスクコントローラは、前記磁気ディスクアレイ装置の内部に前記処理オフロード機能モジュールの存在有無と利用可能なオフロード処理の判別を行うオフロード処理判別部を備える
ことを特徴とする磁気ディスクアレイ装置。
【請求項４】
請求項２に記載の磁気ディスクアレイ装置において、
前記演算処理部は、前記オフロード処理ＦＣコマンドがパリティ演算付きライトであるときに、ＡＴＡ磁気ディスクから読み出すリードコマンド発行と、パリティ演算の実行と、ＡＴＡ磁気ディスクへ書き込むライトコマンド発行と、を行う
ことを特徴とする磁気ディスクアレイ装置。
【請求項５】
請求項２に記載の磁気ディスクアレイ装置において、
前記演算処理部は、前記オフロード処理ＦＣコマンドがオンライン複数ディスクベリファイであるときに、前記コマンド解析処理部から前記オフロード処理ＦＣコマンドを引き継いで、リストとして渡されたＩＤに対応する磁気ディスクにリードコマンドを同時発行する
ことを特徴とする磁気ディスクアレイ装置。
【請求項６】
ＡＴＡ磁気ディスクと、前記ＡＴＡ磁気ディスクを制御するディスクアレイコントローラと、の間のパス上に存在するインターフェースカード内の処理オフロード機能モジュールであって、
前記処理オフロード機能モジュールは、リード又はライトを一例とする標準処理を行う標準処理ＦＣコマンドと、ベンダユニークなオフロード処理を行うオフロード処理ＦＣコマンドと、を前記ディスクコントローラから受け取り、
前記標準処理ＦＣコマンドに対しコマンドマッピングテーブルを用いて対応するＡＴＡコマンドを前記ＡＴＡ磁気ディスクに発行し、前記オフロード処理ＦＣコマンドに対しＡ

ＴＡプロトコルで最適処理となるＡＴＡコマンド群を用意し必要なときには演算する
ことを特徴とする処理オフロード機能モジュール。
【請求項７】
請求項６に記載の処理オフロード機能モジュールにおいて、
前記処理オフロード機能モジュールは、コマンド解析処理部と演算処理部を備え、
前記コマンド解析処理部において前記ディスクアレイコントローラからの標準処理ＦＣコマンドがＡＴＡコマンドにマッピング可能であるか否かを判断し、
前記演算処理部において前記オフロード処理ＦＣコマンドを前記コマンド解析部から引き継いで前記ＡＴＡコマンド群を用意する
ことを特徴とする処理オフロード機能モジュール。
【請求項８】
請求項７に記載の処理オフロード機能モジュールにおいて、
前記オフロード処理ＦＣコマンドは、パリティ演算付きライトコマンド、オンライン複数ディスクベリファイコマンド、ＲＡＩＤフォーマットコマンド、又はディスク間コピーコマンドである
ことを特徴とする処理オフロード機能モジュール。
【請求項９】
ＡＴＡ磁気ディスクと、前記ＡＴＡ磁気ディスクを制御するディスクアレイコントローラと、前記ディスクアレイコントローラと前記ＡＴＡ磁気ディスクとのパス上に存在するインターフェースカードと、を備えた磁気ディスクアレイ装置であって、
前記ＡＴＡ磁気ディスク毎に処理オフロード機能モジュールを接続したディスク収納容器を１つ以上設け、
前記ディスクコントローラは、リード又はライトを一例とする標準処理を行う標準処理ＦＣコマンドと、ベンダユニークなオフロード処理を行うオフロード処理ＦＣコマンドと、を前記インターフェースカードに出力し、
前記処理オフロード機能モジュールは、前記標準処理ＦＣコマンドに対しコマンドマッピングテーブルを用いて対応するＡＴＡコマンドを前記ＡＴＡ磁気ディスクに発行し、前記オフロード処理ＦＣコマンドに対しＡＴＡプロトコルで最適処理となるＡＴＡコマンド群を用意する
ことを特徴とする磁気ディスクアレイ装置。
【請求項１０】
請求項９に記載の磁気ディスクアレイ装置において、
前記処理オフロード機能モジュールは、コマンド解析処理部と演算処理部を備え、
前記コマンド解析処理部において、前記ディスクアレイコントローラからの標準処理ＦＣコマンドがＡＴＡコマンドにマッピング可能であるか否かを判断し、
前記演算処理部において、前記オフロード処理ＦＣコマンドを前記コマンド解析部から引き継いで前記ＡＴＡコマンド群を用意する
ことを特徴とする磁気ディスクアレイ装置。
【請求項１１】
請求項９又は１０に記載の磁気ディスクアレイ装置において、
前記オフロード処理ＦＣコマンドは、パリティ演算付きライトコマンド、オンライン複数ディスクベリファイコマンド、ＲＡＩＤフォーマットコマンド、又はディスク間コピーコマンドである
ことを特徴とする磁気ディスクアレイ装置。
【請求項１２】
磁気ディスクと、前記磁気ディスクを制御するディスクアレイコントローラと、前記ディスクアレイコントローラと前記磁気ディスクとのパス上に存在するインターフェースカード内の処理オフロード機能モジュールと、を備えた磁気ディスクアレイ装置であって、
前記磁気ディスクは、ＡＴＡ磁気ディスクとＦＣ磁気ディスクとが混在配置され、
前記ディスクコントローラは、リード又はライトを一例とする標準処理を行う標準処理

ＦＣコマンドと、ベンダユニークなオフロード処理を行うオフロード処理ＦＣコマンドと、を前記インターフェースカードに出力し、

前記インターフェースカード内の前記処理オフロード機能モジュールは、前記ディスクコントローラからのＦＣコマンドが、前記標準処理ＦＣコマンドであり且つＦＣ磁気ディスクを対象としたコマンドであれば、当該ＦＣコマンドに対して処理を加えずに素通りさせて該当するＦＣ磁気ディスクに送り、前記標準処理ＦＣコマンドであり且つＡＴＡ磁気ディスクを対象としたコマンドであれば、当該ＦＣコマンドに対してコマンドマッピングテーブルを用いて対応するＡＴＡコマンドを前記ＡＴＡ磁気ディスクに発行し、

前記処理オフロード機能モジュールは、前記ディスクコントローラからのＦＣコマンドが、前記オフロード処理ＦＣコマンドであれば、対象とする磁気ディスクの種別と処理内容に応じた最適なコマンド群に変換し処理を行う

ことを特徴とする磁気ディスクアレイ装置。

【請求項１３】

請求項１２に記載の磁気ディスクアレイ装置において、

前記処理オフロード機能モジュールは、コマンド解析処理部と演算処理部を備え、

前記コマンド解析処理部において、前記ディスクアレイコントローラからのＦＣコマンドが、標準処理ＦＣコマンドであり且つＡＴＡ磁気ディスクを対象としたコマンドであれば、ＡＴＡコマンドにマッピング可能であるか否かを判断し、

前記演算処理部において、前記ディスクコントローラからのＦＣコマンドが前記オフロード処理ＦＣコマンドであれば、前記オフロード処理ＦＣコマンドを前記コマンド解析部から引き継いで最適なコマンド群に変換し処理を行う

ことを特徴とする磁気ディスクアレイ装置。

【請求項１４】

磁気ディスクと、前記磁気ディスクを制御するディスクアレイコントローラと、前記ディスクアレイコントローラと前記磁気ディスクとのパス上に存在するインターフェースカードと、を備えた磁気ディスクアレイ装置であって、

前記磁気ディスクは、ＡＴＡ磁気ディスクとＦＣ磁気ディスクとが混在配置され、

前記ディスクアレイコントローラはコントローラ処理部を有し、前記コントローラ処理部に接続された処理オフロード機能モジュールが前記ディスクアレイコントローラ毎に設置され、

前記コントローラ処理部は、リード又はライトを一例とする標準処理を行う標準処理ＦＣコマンドと、ベンダユニークなオフロード処理を行うオフロード処理ＦＣコマンドと、を前記処理オフロード機能モジュールに出力し、

前記ディスクアレイコントローラ内の前記処理オフロード機能モジュールは、前記コントローラ処理部からのＦＣコマンドが、前記標準処理ＦＣコマンドであり且つＦＣ磁気ディスクを対象としたコマンドであれば、当該ＦＣコマンドに対して処理を加えずに素通りさせて該当するＦＣ磁気ディスクに前記インターフェースカードを介して送り、前記標準処理ＦＣコマンドであり且つＡＴＡ磁気ディスクを対象としたコマンドであれば、当該ＦＣコマンドに対してコマンドマッピングテーブルを用いて対応するＡＴＡコマンドを前記ＡＴＡ磁気ディスクに前記インターフェースカードを介して発行し、

前記処理オフロード機能モジュールは、前記コントローラ処理部からのＦＣコマンドが、前記オフロード処理ＦＣコマンドであれば、対象とする磁気ディスクの種別と処理内容に応じた最適なコマンド群に変換し処理を行う

ことを特徴とする磁気ディスクアレイ装置。

【書類名】明細書
【発明の名称】処理オフロード機能モジュールを備えた磁気ディスクアレイ装置
【技術分野】
【０００１】
本発明は、主としてコンピュータの外部記憶装置システムに関わり、特に、磁気ディスクアレイ装置の性能向上を図る技術に関する。
【背景技術】
【０００２】
従来技術として、複数のディスクにデータの処理を分散させ且つ冗長データを記録することによって、高信頼で高速にデータを記録する磁気ディスクアレイ装置（ＲＡＩＤ装置：Redundant Arrays of Inexpensive Disks）が広く用いられている。磁気ディスクアレイの詳細については、例えば非特許文献１に記載されている。
【０００３】
一般に、磁気ディスクアレイ装置は、多数のディスクを搭載しており、これらの磁気ディスクに適切にデータを分散格納し処理するために、入出力制御装置であるディスクアレイコントローラが搭載されている。
【０００４】
ディスクアレイコントローラは、データを適切なディスクに格納するための処理を行う機能の他に、複数のディスクをＲＡＩＤ向けにフォーマットする機能や、ディスクの万一の故障においてもデータを失うことのないように冗長データを計算して書き込む機能や、また、故障予兆のあるディスクを未然に検出する予防保守機能などを有していて、磁気ディスクへの入出力の他にも多数の制御処理を行っている。このため、ディスクアレイコントローラには高速にデータを処理する性能が求められており、改良が進められてきた。
【０００５】
一方で、磁気ディスクアレイ装置の高性能化を図る別の手法として、全ての処理をディスクアレイコントローラに一任するのではなく、配下に処理の一部を肩代わりさせるサブコントローラを設け、これらのサブコントローラに処理を任せるオフロード処理（ディスクコントローラからロード（負荷）をオフにする処理）という方法がある。
【０００６】
オフロード処理により処理高速化を図る方法としては、例えば、特許文献１において、ファイバチャネルループを複数のループに分断し、ループごとに置かれるサブコントローラを用いることで、複数のフォーマットを多重実行する方法が開示されている。
【０００７】
この特許文献１には、データの流れるファイバチャネルループを分断することによって、ループ上を流れるデータを局所化し、更に、コントローラのデータバスの使用効率を改善することが可能である旨が開示されている。また、コントローラが行う処理の一部をサブコントローラに行わせることによって、コントローラのＩ／Ｏ処理負荷を軽減することが可能である旨が開示されている。さらに、ディスク間のコピーや、装置を停止させずにオンライン中にディスクの表面チェックを行う、予防保守の一種であるオンラインベリファイなどの処理も可能である旨が開示されている。
【特許文献１】特開２００１－２２５２６号公報
【非特許文献１】「A Case for Redundant Arrays of Inexpensive Disks (RAID)」, David A.Patterson, Garth Gibson, and Randy H.Katz, CoMPUter Science Division Department of Electrical Engineering and CoMPUter Sciences, University of California Berkeley
【発明の開示】
【発明が解決しようとする課題】
【０００８】
磁気ディスクアレイ装置に対する要望としては、磁気ディスクアレイ装置の高性能化が求められる一方で、データの爆発的増加を背景に業務用途やデータの種類・重要度に応じ

て、格納先となる磁気ディスクに従来のような高いサーバ用ディスクではなくて、より安価なディスクに保存したいという利用者のニーズが生まれてきている。このようなニーズに基づいて、近年では主に個人向けパソコン（ＰＣ）などで多用されてきたＡＴＡ（ＡＴＡｔｔａｃｈｅｍｅｎｔ）規格磁気ディスク（以下、ＡＴＡディスクと称する）を、企業向け磁気ディスクアレイ装置でも利用する動きがある。

【０００９】

ＡＴＡディスクは、ホストのインターフェースコントローラ部に当たる機能をディスク内に内蔵し、従来から主にサーバ用途で多用されてきたＳＣＳＩディスクやＦＣ（Ｆｉｂｒｅ　Ｃｈａｎｎｅｌ）ディスクに比べて、複雑な回路を必要とせず、安価に製造が可能なディスクである。

【００１０】

また、ＡＴＡ規格の論理仕様である制御コマンドは、ディスクコントローラが複雑な判断をせずに済むように、機能の微妙な違いごとに異なるコマンドが割り当てられている。例えば、ライト系コマンドでは、一般のＷＲＩＴＥコマンドのほか、ＤＭＡ（Ｄｉｒｅｃｔ　Ｍｅｍｏｒｙ　Ａｃｃｅｓｓ）が設定されている場合にはＤＭＡ転送でライト行うＷＲＩＴＥ　ＤＭＡ、セクタ単位での書き込みを行うＷＲＩＴＥ　ＳＥＣＴＯＲＳなどがあり、さらに、それぞれのコマンドに、ＷＲＩＴＥ（ＲＥＴＲＹ）、ＷＲＩＴＥ（ＮＯＲＥＴＲＹ）のように失敗時にリトライ処理を行うか否かで２種類の別々のコマンドが用意される。

【００１１】

一方、物理的な仕様においては、元来多数のディスクを接続する必要がないＰＣ内蔵用として設計されているため、パラレルケーブルを用いて接続するＰａｒａｌｌｅｌ　ＡＴＡディスクでは１バスあたりに最大２台しか接続できず、インターフェース高速化のためＰａｒａｌｌｅｌ　ＡＴＡに代わって登場してきたＳｅｒｉａｌ　ＡＴＡ（以下、ＳＡＴＡと称する）では、Ｐｏｉｎｔ　ｔｏ　Ｐｏｉｎｔ接続形態により１バスに１台しか接続できない。また、コントローラとディスク間のケーブルも１ｍ以下と制限がある。このようなことから、多数の磁気ディスクを複数の磁気ディスク格納筐体にわたって格納し、さらに、これらの磁気ディスクをディスクアレイコントローラと接続する必要がある磁気ディスクアレイ装置では、接続形態ならびに接続距離においてＡＴＡディスクをそのまま適用するのは難しい。そのため、磁気ディスクアレイ装置では従来より使われてきたファイバチャネル（以下、ＦＣと称する）にコマンドや物理インターフェースを変換して接続するＦＣ－ＡＴＡ変換接続方式が用いられている。

【００１２】

このＦＣ－ＡＴＡ変換接続方式は、ＡＴＡコマンドを対応するＦＣコマンドにマッピングすることでＦＣ－ＡＴＡ論理変換を実現し、さらに、ＡＴＡとＦＣの信号レベルなどの物理インターフェースを変換することで、ディスクアレイコントローラからはＡＴＡディスクをＦＣディスクであるかのように扱えるようにするものである。

【００１３】

ＦＣ－ＡＴＡ論理変換部においては、２つのプロトコル間で類似するコマンドをマッピングすることとなるが、２つの規格はそれぞれ別の団体が策定した規格であり、まったく異なる仕様を持つ。このため、備える機能もおのずと違い、実際にはＦＣディスクの備える機能に対して、ＡＴＡディスクには該当する機能がないものもある。このような機能を呼び出すＦＣコマンドに対しては、ＡＴＡコマンドをマッピングすることはできない。

【００１４】

従って、ディスクアレイコントローラでは、コマンド発行に先立ち、まずターゲットがＦＣ－ＡＴＡ変換によって、ＦＣコマンドを受け付けられるＡＴＡディスクなのか、単なるＦＣディスクなのかを判別する必要がある。その後、ターゲットがＡＴＡディスクであった場合、発行しようとするコマンドがＦＣ－ＡＴＡ変換で変換可能なコマンドであるか否かを判別し、さらに、変換不可能なコマンドであった場合には、他のＦＣコマンドの組み合わせで、ほぼ同等の処理を行えるようにするか、もしくは、コマンドの発行を取りや

めるかを選択する必要がある。その結果、ディスクアレイコントローラでは判別のための処理が必要となり、処理に要する負荷が増大し、結果として性能を落としてしまう可能性がある。

【００１５】

一方、マッピングされるＡＴＡコマンドから見た場合、ＦＣのコマンドにマッピングできるのは、ＦＣコマンドに仕様としてあるものだけである。上述したように、ＡＴＡは目的の違いに応じて同種のコマンドが複数存在する。しかし、ＦＣ－ＡＴＡ変換においてディスクアレイコントローラから発行されるＦＣコマンドを択一的にマッピングするしかなく、豊富なＡＴＡのコマンドを十分に活かしきれない。また、ＦＣ規格上に該当する機能がないものは、そもそもマッピングすることができないという課題が生じる。

【００１６】

上述したような背景をもつＡＴＡディスクを用いた磁気ディスクアレイ装置に対して、従来技術のようなオフロード処理の適用を考えた場合に次のような課題が生じる。

【００１７】

第１の課題は、コマンドオーバーヘッドの増加である。従来技術のようにサブコントローラを設けた場合、オフロード処理によってディスクアレイコントローラの処理は分散できるものの、ディスクアレイコントローラからの指示は、サブコントローラでＦＣコマンドでの処理に展開され、次に、ＦＣ－ＡＴＡ変換部でＡＴＡコマンドに変換され、そして、ＡＴＡディスクへと伝播するという冗長な経路を取る。即ち、ディスクアレイコントローラ、サブコントローラ、ＦＣ－ＡＴＡ変換部という構成要素が縦続的に接続される経路を構成している。また、この処理に対する応答も同じ経路をたどるため、結果としてコマンドに対するディスクの応答が悪化してしまうおそれがある。

【００１８】

第２の課題は、ＦＣ－ＡＴＡ変換では必ずしも最適なＡＴＡコマンドが選択されないという課題を依然として解決できない点である。前述の特許文献１によれば、サブコントローラはＦＣループ上にあるため、発行できるコマンドもおのずとＦＣコマンドとなってしまう。このため、ＦＣ－ＡＴＡ変換における最適なＡＴＡコマンドの選択に課題が残る。

【００１９】

第３の課題は、ＩＤリソース消費による接続可能なディスク台数の減少である。サブコントローラがディスクコントローラから指示を直接受けるためにはＦＣループ上のデバイスとして認識できる必要があり、この認識のためにループＩＤを消費する。サブコントローラを多数設置すればそれだけ多重化が可能となるが、この多数設置の分だけＩＤが消費されるため、接続できるディスク台数は減少し、大容量化のニーズに相反することになる。

【００２０】

そこで、本発明の目的は、ＡＴＡディスクの利用を可能としながら、ＦＣ－ＡＴＡコマンド変換によるオーバーヘッドを削減し、最適なＡＴＡコマンドを活用し、さらに、ＩＤ等のリソースを無消費にして、オフロード処理を実現する磁気ディスクアレイ装置を提供することにある。

【課題を解決するための手段】

【００２１】

前記課題を解決するために、本発明は主として次のような構成を採用する。

ＡＴＡ磁気ディスクと、前記ＡＴＡ磁気ディスクを制御するディスクアレイコントローラと、前記ディスクアレイコントローラと前記ＡＴＡ磁気ディスクとのパス上に存在するインターフェースカード内の処理オフロード機能モジュールと、を備えた磁気ディスクアレイ装置であって、

前記ディスクコントローラは、リード又はライトを一例とする標準処理を行う標準処理ＦＣコマンドと、ベンダユニークなオフロード処理を行うオフロード処理ＦＣコマンドと、を前記インターフェースカードに出力し、

前記インターフェースカード内の前記処理オフロード機能モジュールは、前記標準処理

ＦＣコマンドに対しコマンドマッピングテーブルを用いて対応するＡＴＡコマンドを前記ＡＴＡ磁気ディスクに発行し、前記オフロード処理ＦＣコマンドに対しＡＴＡプロトコルで最適処理となるＡＴＡコマンド群を用意し必要なときには演算する構成とする。

【発明の効果】

【００２２】

本発明によれば、ＡＴＡディスクの利用を可能としながらも、コマンド変換によるオーバーヘッドの増加を最小限にし、かつ、ディスクの論理インターフェースに合わせた最適なコマンド変換を実現し、ＩＤ等のリソースを無消費にしてオフロード処理を可能とすることができる。

【００２３】

さらに、論理インターフェースの異なる複数種の磁気ディスクを同一の上位コマンドで操作することが可能である。

【００２４】

また、全てのオフロード処理用コマンドは、対象となる磁気ディスクに対して通常（標準）コマンドと同じように発行する形式をとることから、たとえば受け取ったデータを暗号化又は復号化してリード又はライトを行うオフロード処理機能を用意し、すべてのリード又はライトをこの機能を呼び出すベンダユニークコマンドによって行えば、容易に暗号化機能付きディスクアレイ装置を提供することも可能である。

【発明を実施するための最良の形態】

【００２５】

「第１の実施形態」

図１は、本発明の第１の実施形態に係る磁気ディスクアレイ装置における全体構成のブロック図であり、処理オフロード機能モジュールを磁気ディスク格納筐体のインターフェースカードに実装した構成を示す図である。図面において、以下、ｘは任意の整数を表す。

【００２６】

磁気ディスクアレイ装置１０１（以下、ディスクアレイと称する）は、ＳＡＮクライアント１２０ｘからＳＡＮ（Ｓｔｏｒａｇｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）１１１と呼ばれるネットワークを通じてアクセスされる二次記憶装置である。ＳＡＮクライアント１２０ｘはＳＡＮ１１１を経由せず直接接続することも可能である。磁気ディスクアレイ装置１０１は、ＡＴＡ磁気ディスク１７ｘｘに対しデータ転送制御を行うディスクアレイコントローラ１０２ｘと、複数のＡＴＡ磁気ディスクを格納する磁気ディスク格納筐体１０４ｘと、を備えている。ディスクアレイ格納筐体１０４ｘとディスクアレイコントローラ１０２ｘは、ファイバチャネル（以下、ＦＣと称する）１１３ｘ，１１４ｘで接続される。

【００２７】

磁気ディスク格納筐体１０４ｘは、複数のＡＴＡ磁気ディスク１７ｘｘとインターフェースカード１０５ｘとから構成されており、ディスクアレイコントローラ１０２ｘは前記インターフェースカード１０５ｘを通じて磁気ディスク１７ｘｘにデータ転送を行う。インターフェースカード１７ｘｘは、ディスクアレイコントローラ１０２ｘと後続する磁気ディスク格納筐体とを相互に接続するＦＣ接続部１１５（例えば、スイッチなど）と、図示するようにＦＣ接続部１１５に接続される処理オフロード機能モジュール１０６と、から構成される。

【００２８】

尚、図１では、ディスクアレイコントローラは磁気ディスク格納筐体の外部に配置されているが、本発明においては、ディスクアレイコントローラが磁気ディスク格納筐体内にあっても構わない。また、ディスクアレイコントローラにおいては図示の例では２つ記載されているが、１つでも又は３つ以上存在しても構わない。

【００２９】

次に、図２はディスクアレイコントローラ１０２ｘの内部構成を示す図である。フロントエンド接続インターフェースコントローラ２０１はＦＣケーブル１０８ｘを通じ、ＳＡ

Ｎ１１１と接続され、データ転送やコマンドデータの送受信を行う。バックエンド接続インターフェースコントローラ２０７ｘはＡＴＡ磁気ディスク間とデータ送受信を行うためのものである。２０５は中央処理制御部であり、メモリ２０６に格納されたディスクアレイ制御プログラム２０６１を動作させる。メモリ２０６は前記ディスクアレイ制御プログラム２０６１の他に、ディスク情報取得プログラム２０６２とディスク情報管理テーブル２０６３とを備える。ディスク情報取得プログラム２０６２は起動時に接続されたＡＴＡ磁気ディスク１７ｘｘの情報を取得するためのものであり、内部に処理オフロード機能モジュール（以下、機能モジュールと称する）１０６の有無と利用可能なオフロード処理の判別を行うオフロード処理判別部１０３ｘを備える。ディスクアレイ制御プログラム２０６１はオフロード処理判別部１０３ｘに加えて、オフロード処理用のベンダユニークコマンド発行部２０６４を有する。

【００３０】

２０２はデータ転送コントローラであり、ディスクアレイ制御プログラム２０６１やディスク情報取得プログラム２０６２の指示により、フロントエンド接続インターフェースコントローラ２０１、バックエンド接続インターフェースコントローラ２０７ｘ、データバッファ２０３間のデータ移動を行う。

【００３１】

管理ネットワークコントローラ２０４は、ディスクアレイコントローラ１０２ｘの管理用ＬＡＮ１０９ｘを通じ、管理用端末１１０から各種監視やディスク構成や設定の変更を行うために用いられる。管理用端末は管理用ＬＡＮ１０９ｘを経由せず直接接続されていてもよい。

【００３２】

次に、図３は本発明の第１の実施形態に関する処理オフロード機能モジュールの内部構成を示す図である。３０２ｘはＦＣ接続インターフェースであり、２０３ｘを介してＦＣ接続部１１５と接続される。ＦＣインターフェースコントローラ３０３は、ＦＣ接続インターフェース３０２ｘを通じてデータバッファ３０１を用いてデータの送受信を行う。同様に、３０８ｘはＡＴＡ接続インターフェースであり、ＡＴＡ接続ライン３１０ｘを用いてＡＴＡ磁気ディスクドライブ１７ｘｘと接続され、ＡＴＡインターフェースコントローラ３０６はデータバッファ３０１とＡＴＡ接続インターフェース３０８ｘ間のデータ送受信を行う。

【００３３】

ＭＰＵ３０５は、メモリ３０７に格納されたプログラム等を用いて演算処理やコマンド変換、インターフェースコントローラ３０３，３０６の制御を行う処理演算装置である。３０９１はＦＣのＩＤとそれに対応するＡＴＡ磁気ディスクの関連付けを管理するためのディスクＩＤ情報管理テーブルである。ＡＴＡ磁気ディスクはＩＤという概念はなく、直接にＦＣ　ＩＤを持つことができない。そこで、このテーブル３０９１において、ＦＣのＩＤとＡＴＡ接続インターフェースＮｏとの対応関係を記録することにより、間接的にＦＣ　ＩＤの割り当てを実現する。また同様に、ＡＴＡディスクはＦＣディスクが持つディスク固有の識別子であるＷＷＮをもてないため、ＷＷＮとＡＴＡ磁気ディスクの固有番号である製造番号との対応付けについてもこのテーブル３０９１で管理する。

【００３４】

３０９２はオフロード処理時の各コマンドの実行状況を管理するコマンド実行状況管理テーブルであり、コマンドとそのコマンドの実行ステータス、ならびにデータの受信があった場合にはそのバッファのアドレスが記録される。３０９３はコマンド解析処理プログラム、３０９４はステータス処理プログラム、３０９５はＡＴＡコマンドマッピングテーブルであり、３０９６は演算処理プログラムである。

【００３５】

次に、図４は図２に示すディスク情報管理テーブル２０６３の詳細構造を示す図である。４０１は磁気ディスクアレイ装置１０１で管理される磁気ディスクの番号、４０２は磁気ディスクの製造社名（ベンダ名）であり、４０３は製品名、４０４は各磁気ディスクに

固有の製造番号である。４０５はその磁気ディスクの容量を表す。また、４０７は機能モジュール１０６によって提供される処理オフロード機能であり、各磁気ディスクに対してどの機能が利用可能であるかを表す。図示するオフロード機能としては、パリティ演算付きライト、ＲＡＩＤフォーマット、複数ディスクベリファイ、ディスクコピーの機能がある。なお、情報の取得の仕方については後述する。このディスク情報管理テーブルについては、図示するよりも、より多くの項目があってもよいし、項目の順序や名称、記載方法に関して異なっていてもよい。

【００３６】

次に、本発明の第１の実施形態に係る磁気ディスクアレイ装置におけるディスク情報取得処理の動作について、図５を参照しながら以下説明する。図５は、ディスクアレイコントローラ１０２ｘのディスク情報取得プログラム２０６２が、当該コントローラ１０２ｘで使用するＡＴＡ磁気ディスク１７ｘｘを認識し、図４に示したディスク情報管理テーブル２０６３に情報登録を行う動作を示したフローチャートである。破線枠５０１で囲まれた処理はディスクアレイコントローラ１０２ｘでの動作を示し、同様に、５０２は処理オフロード機能モジュール（機能モジュール）１０６での動作を示し、５０３は磁気ディスクでの動作を示す。

【００３７】

ディスクアレイコントローラ１０２ｘは磁気ディスクアレイ装置１０１の電源が投入され、コントローラ内部の初期設定が終了すると、ディスク情報取得プログラム２０６２を起動する（５０００）。

【００３８】

ディスク情報取得プログラム２０６２は、接続されている全ディスクの状態を取得するため、ディスクの構成情報をチェックするＦＣコマンドであるＩｎｑｕｉｒｙコマンドと、磁気ディスクを一意に特定可能なＦＣ　ＩＤを指定して、データバッファ２０３に用意する。なお、ここでは、ＦＣ　ＩＤは０から始まるものとする（５００１、５００２、５００３）。データバッファ２０３にデータが用意できると、当該プログラム２０６２はデータ転送コントローラ２０２（図２を参照）とバックエンド接続インターフェースコントローラ２０７ｘに指示を出し（５００３）、１１３ｘないし１１４ｘを通じて、コマンドを発行する（５００４）。

【００３９】

発行されたコマンドは、各磁気ディスク格納筐体１０４ｘのインターフェースカード１０５ｘを介して伝達する。ディスクアレイコントローラ１０３ｘは該当ＩＤをもつディスクに向かってコマンドを発行しているが、実際にはＡＴＡディスクがＦＣ　ＩＤを直接持つことはできない。そのため、機能モジュール１０６がそのＩＤとディスクの対応付けを管理しており、該当ＩＤを持つ磁気ディスクを管理する機能モジュール１０６がディスクに代わってコマンドを受領する（５００６）

機能モジュール１０６がコマンドを受領すると、機能モジュール１０６内にあるＦＣインターフェースコントローラ３０３は、モジュール内データバッファ３０１にそのコマンドを格納し（５００７）、割り込みを用いてＭＰＵ３０５に通知する（５００８）。ＭＰＵ３０５に通知が来ると、コマンド解析処理プログラム３０９３はモジュール内データバッファ３０１からコマンドを読み込み、解析を行う（５００９）。解析により当該プログラム３０９３はコマンドがＡＴＡのコマンドにマッピング可能かどうか判断する（５０１０）。

【００４０】

マッピング可能な場合、コマンド解析処理プログラム３０９３は、コマンドマッピングテーブル３０９５を用いて対応するＡＴＡコマンドを選択し、この選択したコマンドと、ディスクＩＤ情報管理テーブル３０９１からＦＣ　ＩＤと対応によって得られるＡＴＡ接続インターフェースＮｏ（ＡＴＡ接続インターフェースＮｏとＡＴＡディスクＮｏとは対応付けされている）と、必要であればコマンドの実行により得られるデータを格納するモジュール内データバッファ３０１の格納先アドレスの３つを用意し、前記データバッファ

３０１内に用意する（５０１１）。もし、マッピング可能でなかった場合には、コマンド解析処理プログラム３０９３は演算処理プログラム３０９６に処理を移し、適切な処理を行う（５０１２）。なお、演算処理プログラム３０９６に処理を移した場合の詳細な動作については別の処理において説明する。

【００４１】

コマンドが前記データバッファ３０１に用意できると、コマンド解析処理プログラム３０９３はＡＴＡインターフェースコントローラ３０６に指示をして、モジュール内データバッファ３０１で指定されたＮｏをもつＡＴＡ接続インターフェースを介してＡＴＡ磁気ディスクにコマンドを発行する（５０１４）。

【００４２】

コマンドに対しディスクが応答すると（５０１６）、応答はＡＴＡ接続インターフェース３０８ｘを介して、ＡＴＡインターフェースコントローラ３０６によってデータバッファ３０１に格納され（５０１８）、格納したことをＭＰＵ３０５に通知する（５０１９）。

【００４３】

以下、上述した手順とは逆に、ステータス処理プログラム３０９４は該当する応答内容から、ＦＣコマンド用のステータスを作成し（５０２０）、バッファに格納（５０２１）し、ＦＣインターフェースコントローラ３０３に指示を出し（５０２２）、ステータスをディスクアレイコントローラ１０３ｘに報告する（５０２３）。以上の一連の動作により、ディスクアレイコントローラ１０３ｘはコマンドの結果を得るが、ディスクアレイコントローラ１０３ｘからは直接磁気ディスクが応答したかのように見えるため、処理過程において、処理オフロード機能モジュールの仲介はディスクアレイコントローラ１０３ｘにとって透過的である。

【００４４】

報告を受けた磁気ディスクアレイ装置は、同様の手順によって、ＲＥＡＤ　ＣＡＰＡＣＩＴＹコマンドでディスクの容量を取得する（５０２５）。また、ベンダユニークなコマンドによって、該当ディスクに対し利用可能なオフロード処理機能についての情報を問い合わせ（５０２６）、得られた情報をディスク情報管理テーブル２０６３に登録していく（５０２７）。なお、ベンダユニークコマンドの処理に関しては後述する。１台のディスクが終わると、最後のＩＤまで同様の処理を繰り返す（５０２８，５０２９）。

【００４５】

以上のようにしてディスク情報を登録するが、全てのディスクに対して情報を取得できるのであれば、他の方法でも良い。また、１コマンドで機能モジュールに接続される全ディスクの情報を取得するようなベンダユニークコマンドを用意し、そのベンダユニークコマンドによって情報取得してもかまわない。また、図４のディスク情報管理テーブル２０６３に示した管理情報は一例であり、項目内容や情報や登録表記方法が異なっていても構わない。

【００４６】

次に、本発明の第１の実施形態に係る磁気ディスクアレイ装置において、複数のディスクに異なるコマンドを送るオフロード処理について以下説明する。複数のディスクに異なるコマンドを送るオフロード処理例として、パリティ演算付きライトの動作を説明する。図６は、図５に示す処理５０１０において演算処理プログラムが呼び出された場合の処理５０１２を示したものである。

【００４７】

ディスクアレイコントローラ１０２ｘは、ホスト１２０ｘからコマンドとライトするデータを受け取ると、ディスクアレイ制御プログラム２０６１における処理の過程上でパリティ演算が必要かどうかを判断する。演算が必要であった場合、ディスクアレイ制御プログラム２０６１はディスク情報管理テーブル２０６３を参照し、データを書き出す磁気ディスク１７ｘｘが該当オフロード処理を実行可能かどうか判別する。判別の結果、実行可能な場合は、当該処理を指示するベンダユニークなコマンドを生成し、図５に示すステッ

プ５００３～５００５の動作によりライトを行うディスクのうちの１つに対して発行する。コマンドは、コマンド識別番号、データを書き出すディスクのＩＤとパリティを書き出すディスクのＩＤからなるＩＤリスト、ライトするデータの長さ、などの情報を持つ。なお、リストにおいてデータ書き出すディスクのＩＤは複数あってよい。

【００４８】

コマンドを発行されたディスクを管理する機能モジュール１０６がコマンドを受領すると、前述した図５のステップ５００７の動作によりコマンドとライトする新データをモジュール内データバッファ３０１に格納する。その後、ステップ５００９の動作により、コマンド解析処理プログラム３０９３はコマンド解析を開始する。このコマンド３０９３はオフロード処理を行うベンダユニークコマンドであるため、図５に示すステップ５０１０、５０１２の動作により、その処理を演算処理プログラム３０９６に引き継ぐ。

【００４９】

演算処理プログラム３０９６の目的は、与えられたオフロード処理を達成するためにＡＴＡプロトコルで最適な処理となるようコマンド群を用意し、必要であれば演算を行い、目的の処理を達成し、結果をディスクアレイコントローラ１０２ｘに報告することである。このコマンド３０９６で行う処理は、新データと旧データ並びに旧パリティデータから演算によって新パリティデータを算出し、該当ディスクへ新データ、新パリティデータを書き出すことである。この処理を大きく分けると、旧データ及び旧パリティの読み出し（処理Ａ）、新パリティの算出（処理Ｂ）、新データ及び新パリティの書き込み（処理Ｃ）の大きく３段階に分かれる。

【００５０】

演算処理プログラム３０９６は、まず、旧データ及び旧パリティの読み出し処理（処理Ａ）を実現するために、コマンド解析処理プログラム３０９３が解析した結果から、データを書き出すディスクのＩＤ群とパリティを書き出すディスクのＩＤを取得し、ディスクＩＤ情報管理テーブル３０９１より、対応するＡＴＡ接続インターフェースＮｏを順に取得する（６００１）。次に、処理に最適なＡＴＡリードコマンドを選択し（６００２）、このコマンドと、前記ＡＴＡ接続インターフェースＮｏと、リードしたデータを格納するためのモジュール内バッファ３０１のデータ格納アドレスとを、データないしパリティのリードが必要なディスク台数分繰り返して、コマンド実行状況管理テーブル３０９２に登録する（６００３，６００４）。ここで、処理に最適なＡＴＡリードコマンドには、前述した旧データ及び旧パリティの読み出し処理が旧データと旧パリティを高速かつ確実に読み上げることであることから、高速転送可能でかつリードに失敗すると再試行を試みてくれるＲＥＡＤ　ＤＭＡ（ＲＥＴＲＹ）コマンドを用いる。コマンドの登録が終了したらこの管理テーブル３０９２の情報を、モジュール内データバッファ３０１に転送し（６００５）、コマンド実行状況管理テーブル３０９２の各コマンドの処理状況を「コマンド発行待ち」に設定する（６００６）。

【００５１】

モジュール内データバッファ３０１にコマンドの登録が終わると演算処理プログラム３０９６は、ＡＴＡインターフェースコントローラ３０６に指示をして前述した一連のコマンドを発行させる（６００７）。ＡＴＡインターフェースコントローラ３０６はコマンドの発行が終わると、割り込みによって発行処理終了を知らせる（６００８）。演算処理プログラム３０９６は発行終了割り込みを受けると、コマンド実行状況管理テーブル３０９２に登録したコマンドの処理状況を、「応答待ち」に変更する（６００９）。

【００５２】

その後、当該磁気ディスク１７ｘｘがコマンドを受領し、データの転送を終了し、応答ステータスを返すと（６０１１，６０１２）、ＡＴＡインターフェースコントローラ３０６は応答ステータスを受領したことを、割り込みによって演算処理プログラム３０９６に通知する（６０１３）。演算処理プログラム３０９６はモジュール内データバッファ３０１からステータスを読み取り、成功していればコマンド実行状況管理テーブル３０９２における該当コマンドの処理状況を「完了」にし、リードしたデータが格納されているデー

タバッファアドレスをコマンド実行状況管理テーブル３０９２に登録し、失敗していればエラーとして、そのエラー内容を処理状況に登録する（６０１４）。

【００５３】

コマンド実行状況管理テーブル３０９２上のコマンドの処理状況がすべて「完了」になった場合、演算処理プログラムは、前述した３段階の内の新パリティの算出処理に移行する。もし、このテーブル３０９２上に１つでもエラーがある場合には、演算処理プログラムはエラーレポートを作成し、ディスクアレイコントローラ１０２ｘにエラー通知を行い、以降、新パリティの算出処理、新データ及び新パリティの書き込み処理に移らずに、オフロード処理を終了する（６０１７）。エラーレポートは、エラー通知とともにディスクアレイコントローラ１０２ｘに通知される。なお、コマンド実行状況管理テーブル３０９２上のコマンドに「応答待ち」が残っていた場合、一定時間を待って、さらに応答が帰ってこない場合には当該コマンドはタイムアウトとしてエラー処理する（６０１５，６０１６，６０１７）。

【００５４】

３段階の内の新パリティの算出処理（処理Ｂ）に移行すると、演算処理プログラム３０９６は、あらかじめ指定したデータバッファのアドレスに格納されていた新データと旧データ及び旧パリティの読み出し処理によって取得した旧データ、旧パリティデータより排他的演算処理を行い、新パリティデータを算出する（６０１９）。なお、パリティデータは他の方法で算出されてもかまわず、また、図６ではパリティデータは１つとしているが、複数のパリティデータがあってもよい。

【００５５】

次に、３段階の内の新データ及び新パリティの書き込み処理（処理Ｃ）に移行し、最適なＡＴＡライトコマンドを選択し、ライトを行う全ディスクに対してコマンドと該当ディスクが接続されている接続インターフェースＮｏ、書き込みデータが格納されているバッファのアドレスをコマンド実行状況管理テーブル３０９２に登録し（６０２０，６０２１，６０２２）、次に、旧データ及び旧パリティの読み出し処理と同様に、モジュール内データバッファ３０１にコマンドを格納してＡＴＡインターフェースコントローラ３０６にコマンド発行を指示し、ディスクからの応答を待つ（６１００）。パリティ演算でのライトは確実に高速かつ書き込めることが望まれるため、ＡＴＡコマンドにはＷＲＩＴＥ　ＤＭＡ（ＲＥＴＲＹ）を利用する。

【００５６】

演算処理プログラム３０９６はＡＴＡインターフェースコントローラ３０６を通じて新データと新パリティを無事書き込めたという報告を受け取ると、ＦＣステータスを作成し（６０３０）、ディスクアレイコントローラ１０２ｘに報告を行い（６０３３）、オフロード処理を終了する。

【００５７】

次に、本発明の第１の実施形態に係る磁気ディスクアレイ装置において、複数のディスクに対して同一のコマンドを送る処理について以下説明する。複数のディスクに対して同一のコマンドを送る処理例として、オンラインベリファイ処理の動作を説明する。ここで、オンラインベリファイとは、ホストからのＩ／Ｏとは非同期でディスクの表面チェックを行い、故障予兆のあるディスクを検出することを目的とする保守作業である。

【００５８】

通常、ディスクアレイコントローラでオンラインベリファイを行う際は、各ディスクに対しＶＥＲＩＦＹコマンドをディスク全体がチェックし終わるまで複数回発行し、さらにこれをディスク台数分繰り返す必要がある。本発明では、オフロード処理用コマンドとして、コマンド識別番号と同時実行したいディスクのＩＤをリストにしたものを、代表ディスクにのみ１コマンドを発行するだけで、対象とする全てのディスクに対して、同時にオンラインベリファイ処理を実行可能するものである。以下にその動作を説明する。

【００５９】

ディスクアレイコントローラ１０２ｘは、動作経過時間などの或る契機によってオンラ

インベリファイが必要となったディスクがあると、実行対象となるディスクのＩＤをリストにする。ここで対象となるディスクは単数でもよいが、ここでの説明では複数を想定する。ディスクアレイコントローラ１０２ｘはオンラインベリファイ処理の必要な磁気ディスクのＦＣ　ＩＤのリストとコマンド識別番号を用いて、オンラインベリファイオフロード処理実行用のベンダユニークなコマンドを作成し、代表ディスクに発行する。ここで、代表ディスクとはリストに渡されるＩＤを割り当てられたディスクのうちの１つである。なお、ディスクアレイコントローラがコマンドを発行する手順については、図５のステップ５００３～５００５の動作と同じく、データバッファ２０３にコマンドと送信先ＩＤを格納し、データ転送コントローラ２０２と、バックエンド接続インターフェースコントローラ２０７ｘに指示することで行う。

【００６０】

当該ＩＤを管理する機能モジュール１０６がコマンドを受け取ると、コマンド解析処理プログラム３０９３は受け取ったコマンドがオフロード処理用コマンドであると判別し、演算処理プログラム３０９６に処理を渡す。演算処理プログラム３０９６では、リストとして渡されたＩＤに対応するディスクごとに、ベリファイを行うためのコマンドをモジュール内データバッファ３０１に用意する。ベリファイにはディスクのセクタごとに設けられたＥＣＣ情報のみをチェックする方法と、セクタ全体を読み出せるかチェックする方法とがあるが、前者のチェックはセクタに格納されたデータが読み出せるかどうかをチェックの対象とはしないため信頼性は低い。

【００６１】

ここでの説明では、後者の方法によるセクタ全体に対するベリファイを考えるが、セクタ全体をチェックするベリファイコマンドはＡＴＡコマンドに存在しない。そこで、セクタ単位でのリードを試み、かつ失敗した際には再試行を試みずにすぐにエラーを返すＲＥＡＤ　ＳＥＣＴＯＲＳ（ＮＯＲＥＴＲＹ）コマンドによりデータがセクタごとに読み出せるかチェックすることにより同等の機能を実現する。

【００６２】

このコマンドが全ての対象ディスクに対して実行されるように、コマンド実行状況管理テーブル３０９２に登録が終わると、このコマンドリストをモジュール内データバッファ３０１に用意し、ＡＴＡインターフェースコントローラ３０６に指示を出してコマンドを発行する。コマンドに対して全てのディスクから応答が返り、応答の中にエラーがなければ、上述した方法と同様の手順によって、各ディスクの次のセクタを検査する。これをディスクの領域の最後まで繰り返す。エラーが発生した場合には、そのディスクとセクタ位置を特定できるエラーレポートを作成し処理を続行する。

【００６３】

全ての領域の検査が終われば、演算処理プログラム３０９６はディスクアレイコントローラ１０２ｘに報告して、オフロード処理を完了する。

【００６４】

以上説明したように、本発明の第１の実施形態に係る磁気ディスクアレイ装置の概要を図１１を参照して再度説明する。図１１は本発明の実施形態に係る磁気ディスクアレイ装置におけるディスクアレイコントローラ及び処理オフロード機能モジュールの処理態様の概要を説明する図である。

【００６５】

ディスクアレイコントローラは、ホストからのコマンドとデータに基づいて、通常（標準）のコマンド（例えば、リード、ライト）と、ベンダユニーク（製造元独自）なコマンド（例えば、パリティ演算付きライト、ＲＡＩＤフォーマット、オンライン複数ベリファイ、ディスク間コピー（図４参照））とを生成して、処理オフロード機能モジュールに出力する。処理オフロード機能モジュールは、図５に示すように、コマンド解析処理プログラム、演算処理プログラム、及びマッピングテーブルなどを備えており、ディスクアレイコントローラからのＦＣコマンドをコマンド解析処理プログラムによって解析して、ＡＴＡコマンドにマッピング可能か否かを判断し、標準ＦＣコマンドの場合には、マッピング

テーブルから対応するＡＴＡコマンドを選択してＡＴＡディスクに当該ＡＴＡコマンドを発行する。

【００６６】

図１１において、ベンダユニークな４つのコマンドの内で、ＲＡＩＤフォーマットとオンライン複数ベリファイのコマンドは、ディスクアレイコントローラベンダユニークなオフロード処理コマンドであり、パリティ演算付きライトとディスク間コピーのコマンドは、ディスクベンダユニークなオフロード処理コマンドである。なお、ディスクアレイコントローラからのディスクベンダユニークなＦＣコマンド（ディスクアレイコントローラからのＦＣコマンドはＦＣディスクのＩＤを指定している）は、ディスク群にＦＣディスクがあれば、機能モジュールで処理されることなくＦＣディスクにコマンド発行されることとなる。

【００６７】

ベンダユニークなＦＣコマンドはコマンド解析処理プログラムによって１対１のＡＴＡコマンドにマッピングできないと判断されて、このベンダユニークＦＣコマンドは機能モジュール内の演算処理プログラムに引き継がれて、図６に示すフローにしたがって演算処理される。図１１におけるパリティ演算付きライトコマンドの例で説明すると、旧データと旧パリティのＲＥＡＤ（ＲＥＴＲＹ）がＡＴＡディスクに発行され、次に、パリティ演算されて新データと新パリティがＡＴＡディスクにＷＲＩＴＥ（ＲＥＴＲＹ）される。図１１において実線はデータの流れを示し、破線はステータスの流れを示す。

【００６８】

本実施形態において、ＡＴＡディスクにはＩＤが付加されていないので、サブコントローラである処理オフロード機能モジュールがＡＴＡディスクを管理することになる。ディスクアレイコントローラは配下のＩＤに対してコマンドを発行するため、ディスクコントローラからのコマンドは一旦サブコントローラが全て受け取って処理した後にＡＴＡディスクにコマンドを発行することになる。換言すると、ディスクコントローラは配下のディスクＩＤに向かってコマンドを発行しているのであるから、サブコントローラの存在は見えていない（換言すると、サブコントローラはディスクアレイコントローラに対して透過的にディスクを見せる機能を奏する）ことになる。

【００６９】

一方、従来技術においては、ディスクアレイコントローラ配下にサブコントローラとＦＣディスクがＦＣループを形成していて、このＦＣループ上にサブコントローラとＦＣディスクとがそれぞれＩＤを有しているので（ＦＣ規格上で１ループ当たり１２７台のサブコントローラとＦＣディスクが設置可能）、ディスクアレイコントローラはＩＤで管理しているので、サブコントローラとＦＣディスクを別々に意識してこれらにそれぞれコマンド発行する必要があり、場合によってはコマンド実行順序が保証されない（例えば、ＦＣディスクからリードした後にライトすべきところを、先にＦＣディスクにライトするという不都合が生じ得る）。また、従来技術では、サブコントローラとＦＣ－ＡＴＡ変換部とが別個に縦続接続されていたので、サブコントローラでオフロード処理のＦＣコマンドが発せられても、ＦＣ－ＡＴＡ変換部において、入力されたＦＣコマンドに１対１のＡＴＡコマンドが対応しない場合が多々あった。

【００７０】

本発明の実施形態においては、処理オフロード機能モジュールがその内部に、図３に示すように、コマンド解析処理プログラム、演算処理プログラム、ＡＴＡコマンドマッピングテーブル等を備えて、図５と図６の処理フローを実行することで、オフロード処理を内部でＡＴＡコマンドに展開することができて、ＡＴＡコマンドには直接マッピング不可能なＦＣコマンドが変換利用できることとなる。また、本実施形態の機能モジュール（サブコントローラ）がＦＣループＩＤとＡＴＡディスクの対応付けを管理するため、機能モジュールが自分用のＩＤを持つ必要が無く、即ち、ディスクアレイコントローラが管理の対象とするＩＤを、機能モジュールで使用（消費）することがない。また、本実施形態では、ディスクアレイコントローラは、標準コマンドもディスクベンダユニークコマンド（パ

リティ演算付けライト、ディスク間コピー）もコントローラベンダユニークコマンド（ＲＡＩＤフォーマット、オンライン複数ベリファイ）も、サブコントローラとディスクとを意識し区分けしてそれらのコマンドを発行する必要がなく、対象ディスクのＩＤに対して発行すれば済む（機能モジュールが全てのコマンドを受け取ってディスク管理するので）。

【００７１】

このように、本発明の第１の実施形態に係る磁気ディスクアレイ装置によれば、オフロード処理用ＦＣコマンドを直接ＡＴＡコマンドに変換して処理するため、コマンド変換によるオーバーヘッドを最小にして実行することができる。また、オフロード処理の内容に応じて最適なＡＴＡコマンドを用いて処理をすることが可能である。また、本実施形態と同様の方法によって、ドライブ間のデータコピーや複数ディスクにＲＡＩＤフォーマット処理も行うことが可能である。

【００７２】

「第２の実施形態」

図７は、本発明の第２の実施形態に係る磁気ディスクアレイ装置における全体構成のブロック図であり、処理オフロード機能モジュールを磁気ディスク格納筐体のキャニスタに実装した構成を示す図である。図７において、８０ｘはキャニスタ（ディスク収納容器）を表し、８１０ｘは各キャニスタ内に組み込まれた処理オフロード機能モジュールを表す。

【００７３】

図８は処理オフロード機能モジュールをキャニスタ内に実装した場合の内部構成を示す図である。８２００は第２の実施形態に関する処理オフロード機能モジュール８１０ｘを搭載したコネクタボードであり、磁気ディスク格納筐体側接続インターフェース８２０１とディスク側接続インターフェース接続用インターフェース８２０２を備える。

【００７４】

本発明の第２の実施形態によれば、上述した第１の実施形態と同様の効果を得られる上に、ディスクごとにオフロード処理が可能となる。オフロード処理で生じるデータの移動はキャニスタ内に限定されるため、データのバス効率はよりよくなる。また、処理の並列多重度が上がり、ディスクアレイコントローラやアレイコントローラにつながるバス上にほとんど負荷をかけることなく単位時間当たり処理性能を向上させることが可能である。

【００７５】

「第３の実施形態」

図９は、本発明の第３の実施形態に係る磁気ディスクアレイ装置における処理オフロード機能モジュールのブロック図であり、ＦＣディスクとＡＴＡディスクを混載可能とした場合の構成を示す図である。図９において、同一の磁気ディスク格納筐体１０４ｘ内に、収めるＡＴＡディスクの一部をＦＣディスクに置き換えて混在させた場合における処理オフロード機能モジュール１０６の内部構成を示している。

【００７６】

本発明の第１の実施形態の構成に比べて、混在を実現するために、機能モジュール１０６はＦＣディスク９７０ｘを接続するためのＦＣ接続インターフェース９０８ｘと、このインターフェースをコントロールするためのＦＣインターフェースコントローラ９０６と、オフロード処理を適切なＦＣコマンド群に展開するためのＦＣコマンドマッピングテーブル９０９８と、を新たに備えている。

【００７７】

本発明の第１の実施形態との動作上の差異について以下説明する。ＦＣインターフェースコントローラ３０３が上位のＦＣ接続インターフェース３０２ｘを通じてコマンドを受け取ると、コマンド解析処理プログラム３０９３はディスクＩＤ情報管理テーブル３０９１を参照して、そのコマンドがＡＴＡディスクに対するものなのかＦＣディスクに対するものなのかを判別する。送られてきたコマンドが通常のＦＣコマンドであり、かつＦＣディスクを対象としているのであれば、当該ＦＣコマンドを変換することなく素通りさせ、

ＡＴＡディスクが対象であった場合は、ＡＴＡコマンドマッピングテーブル３０９６を用いて、最適なＡＴＡコマンド群に変換する。

【００７８】

また、受け取ったコマンドがオフロード処理用コマンドだった場合、コマンド解析処理プログラム３０９３は演算処理プログラム３０９７に処理を移し、対象ディスクの種別と処理内容に応じて最適なコマンド群に変換して処理を行う。

【００７９】

本発明の第３の実施形態によれば、第１の実施形態で得られる効果に加え、異なるディスクを同一の上位コマンドで操作することが可能であり、また、その上位コマンドをディスクの種類と処理内容に応じた最適なコマンド群を用いて処理することが可能である。

【００８０】

「第４の実施形態」

図１０は、本発明の第４の実施形態に係る磁気ディスクアレイ装置における全体構成のブロック図であり、処理オフロード機能モジュールをディスクアレイコントローラに実装した構成を示す図である。図１０にはディスクアレイコントローラ１０２ｘに処理オフロード機能モジュール１１０３ｘを実装した場合の磁気ディスクアレイ装置の構成示している。

【００８１】

処理オフロード機能モジュール１１０３ｘを搭載するディスクアレイコントローラ１０２ｘは、複数のＡＴＡディスクをスイッチ的に接続するＡＴＡ接続部１１１１０と、同様の方法によってＦＣディスクを接続するＦＣ接続部１１１２０とを介して、ＡＴＡディスク１７ｘｘと、ＦＣディスク１１０ｘｘとに接続される。なお、図１０において、ｘ、ｙ、ｚ、ａ、ｂ、ｍ、ｎは任意の自然数である。

【００８２】

また、図１０では、ＡＴＡディスクとＦＣディスクを挙げているが、これ以外のディスクの組み合わせでもよい。例としては、ＡＴＡディスクと、ＳＡＳ（Ｓｅｒｉａｌ　Ａｔｔａｃｈｅｄ　ＳＣＳＩ）ディスクの組みあわせ等であってもよい。また、磁気ディスクの種類は２種類以上あってもよい。

【００８３】

本発明の第４の実施形態によれば、ディスクアレイコントローラは一種類のコマンドが発行できればよく、下位に接続されるディスクのインターフェースごとにコマンドを用意する必要がなくなり、上述した第３の実施形態と同様に、ディスクに応じた最適なコマンド処理を実行することが可能である。また、処理オフロード機能モジュールはディスクアレイコントローラ上に実装されるため、ディスクアレイコントローラ処理部（図２に示す構成と同様な構成）と処理オフロード機能モジュール間のプロトコルは、ベンダユニークな第３のコマンド体系であってもよい。

【図面の簡単な説明】

【００８４】

【図１】本発明の第１の実施形態に係る磁気ディスクアレイ装置における全体構成のブロック図であり、処理オフロード機能モジュールを磁気ディスク格納筐体のインターフェースカードに実装した構成を示す図である。

【図２】本発明の第１実施形態におけるディスクアレイコントローラの内部構成を示す図である。

【図３】本発明の第１の実施形態に関する処理オフロード機能モジュールの内部構成を示す図である。

【図４】図２に示すディスク情報管理テーブル２０６３の詳細構造を示す図である。

【図５】本発明の第１の実施形態に係る磁気ディスクアレイ装置におけるディスク情報取得処理の動作を示すフローチャートである。

【図６】本発明の第１の実施形態に係る磁気ディスクアレイ装置において、複数のディスクに異なるコマンドを送るオフロード処理例としてのパリティ演算付きライトの

動作を示すフローチャートである。

【図７】本発明の第２の実施形態に係る磁気ディスクアレイ装置における全体構成のブロック図であり、処理オフロード機能モジュールを磁気ディスク格納筐体のキャニスタに実装した構成を示す図である。

【図８】本発明の第２の実施形態に関する処理オフロード機能モジュールをキャニスタ内に実装した場合の内部構成を示す図である。

【図９】本発明の第３の実施形態に係る磁気ディスクアレイ装置における処理オフロード機能モジュールのブロック図であり、ＦＣディスクとＡＴＡディスクを混載可能とした場合の構成を示す図である。

【図１０】本発明の第４の実施形態に係る磁気ディスクアレイ装置における全体構成のブロック図であり、処理オフロード機能モジュールをディスクアレイコントローラに実装した構成を示す図である。

【図１１】本発明の実施形態に係る磁気ディスクアレイ装置におけるディスクアレイコントローラ及び処理オフロード機能モジュールの処理態様の概要を説明する図である。

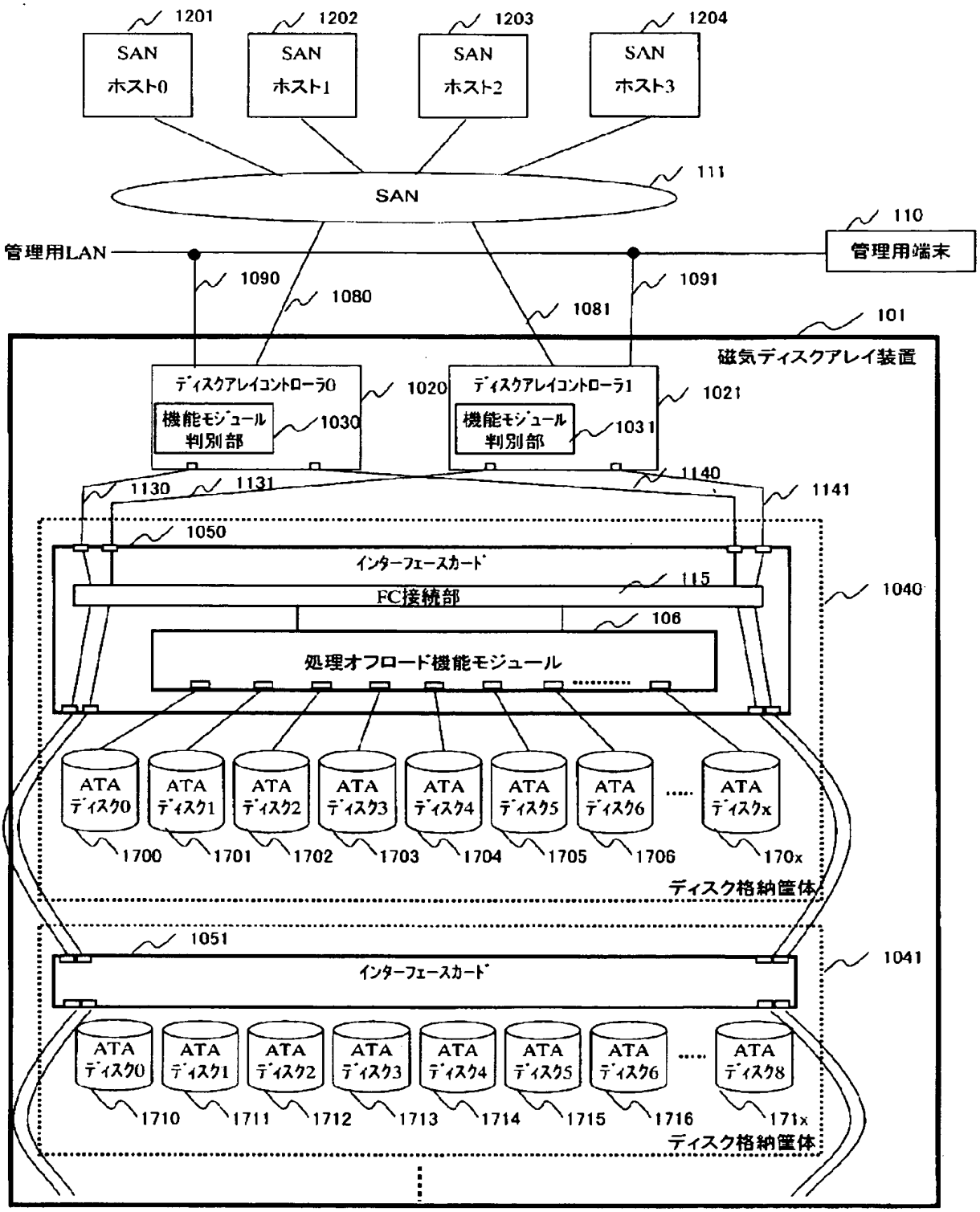
【符号の説明】

【００８５】

１０１　磁気ディスクアレイ装置
１１０　管理用端末
１１１　ＳＡＮ（ストレージエリアネットワーク）
１１５　ＦＣ接続部
１０６　処理オフロード機能モジュール
１２０ｘ　ＳＡＮホスト
１０２ｘ　ディスクアレイコントローラ
１０４ｘ　磁気ディスク格納筐体
１０５ｘ　インターフェースカード
１７０ｘ，１７１ｘ　ＡＴＡ磁気ディスク
２０６１　ディスクアレイ制御プログラム
２０６２　ディスク情報取得プログラム
２０６３　ディスク情報管理テーブル
３０９２　コマンド実行状況管理テーブル
３０９３　コマンド解析処理プログラム
３０９４　ステータス処理プログラム
３０９５　ＡＴＡコマンドマッピングテーブル
３０９６　演算処理プログラム
８０ｘ　キャニスタ
８２００　コネクタボード
３１０ｘ　ＡＴＡ接続ライン
９１０ｘ　ＦＣ接続ライン

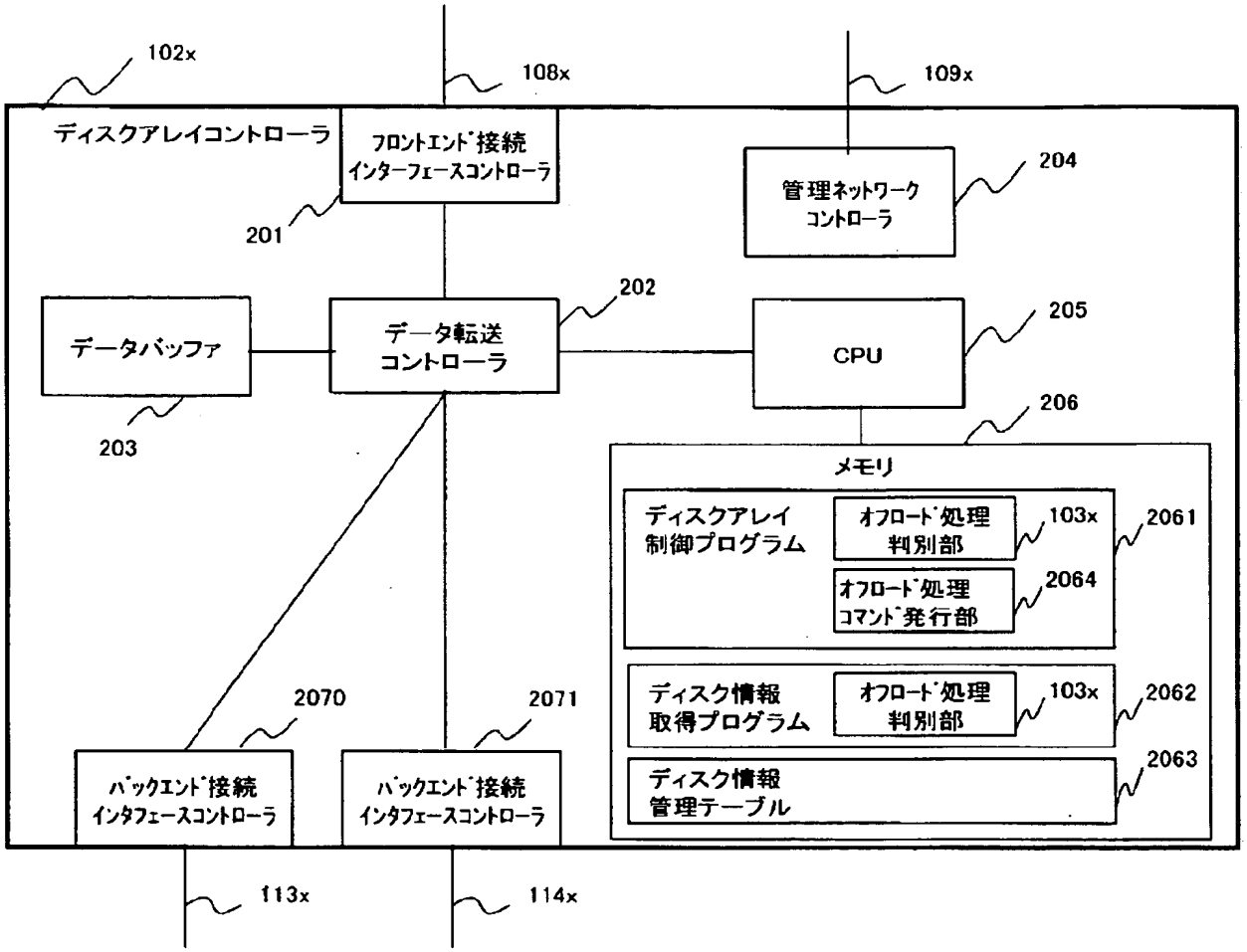
【書類名】図面
【図１】

図1


1201
SAN ホスト0
1202
SAN ホスト1
1203
SAN ホスト2
1204
SAN ホスト3
111
SAN
110
管理用端末
管理用LAN
1090
1080
1081
1091
101
磁気ディスクアレイ装置
ディスクアレイコントローラ0
1020
機能モジュール判別部
1030
ディスクアレイコントローラ1
1021
機能モジュール判別部
1031
1130
1131
1140
1141
1050
インターフェースカード
115
FC接続部
1040
106
処理オフロード機能モジュール
ATA ディスク0
ATA ディスク1
ATA ディスク2
ATA ディスク3
ATA ディスク4
ATA ディスク5
ATA ディスク6
ATA ディスクx
1700
1701
1702
1703
1704
1705
1706
170x
ディスク格納筐体
1051
インターフェースカード
1041
ATA ディスク0
ATA ディスク1
ATA ディスク2
ATA ディスク3
ATA ディスク4
ATA ディスク5
ATA ディスク6
ATA ディスク8
1710
1711
1712
1713
1714
1715
1716
171x
ディスク格納筐体

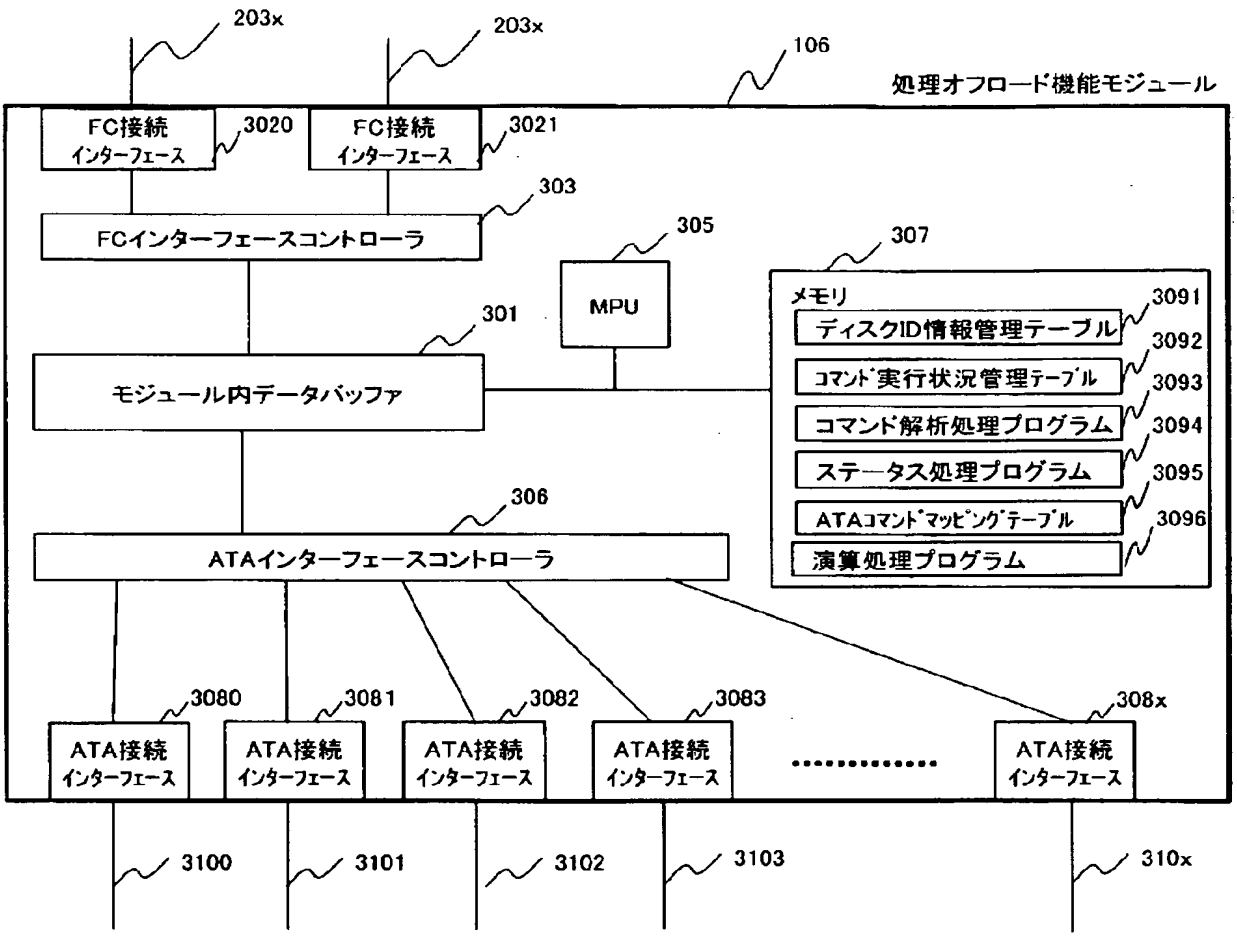
【図２】

図2


102x
108x
109x
ディスクアレイコントローラ
フロントエンド接続
インターフェースコントローラ
201
管理ネットワーク
コントローラ
204
データバッファ
データ転送
コントローラ
202
CPU
205
203
206
メモリ
ディスクアレイ
制御プログラム
オフロード処理
判別部
103x
2061
オフロード処理
コマンド発行部
2064
ディスク情報
取得プログラム
オフロード処理
判別部
103x
2062
ディスク情報
管理テーブル
2063
2070
2071
バックエンド接続
インタフェースコントローラ
バックエンド接続
インタフェースコントローラ
113x
114x

【図３】

図3


203x
203x
106
処理オフロード機能モジュール
FC接続 インターフェース
3020
FC接続 インターフェース
3021
303
FCインターフェースコントローラ
305
MPU
307
メモリ
3091
ディスクID情報管理テーブル
3092
ｺﾏﾝﾄﾞ実行状況管理ﾃｰﾌﾞﾙ
3093
コマンド解析処理プログラム
3094
ステータス処理プログラム
3095
ATAｺﾏﾝﾄﾞﾏｯﾋﾟﾝｸﾞﾃｰﾌﾞﾙ
3096
演算処理プログラム
301
モジュール内データバッファ
306
ATAインターフェースコントローラ
3080
3081
3082
3083
308x
ATA接続 インターフェース
ATA接続 インターフェース
ATA接続 インターフェース
ATA接続 インターフェース
ATA接続 インターフェース
3100
3101
3102
3103
310x

【図４】

図4

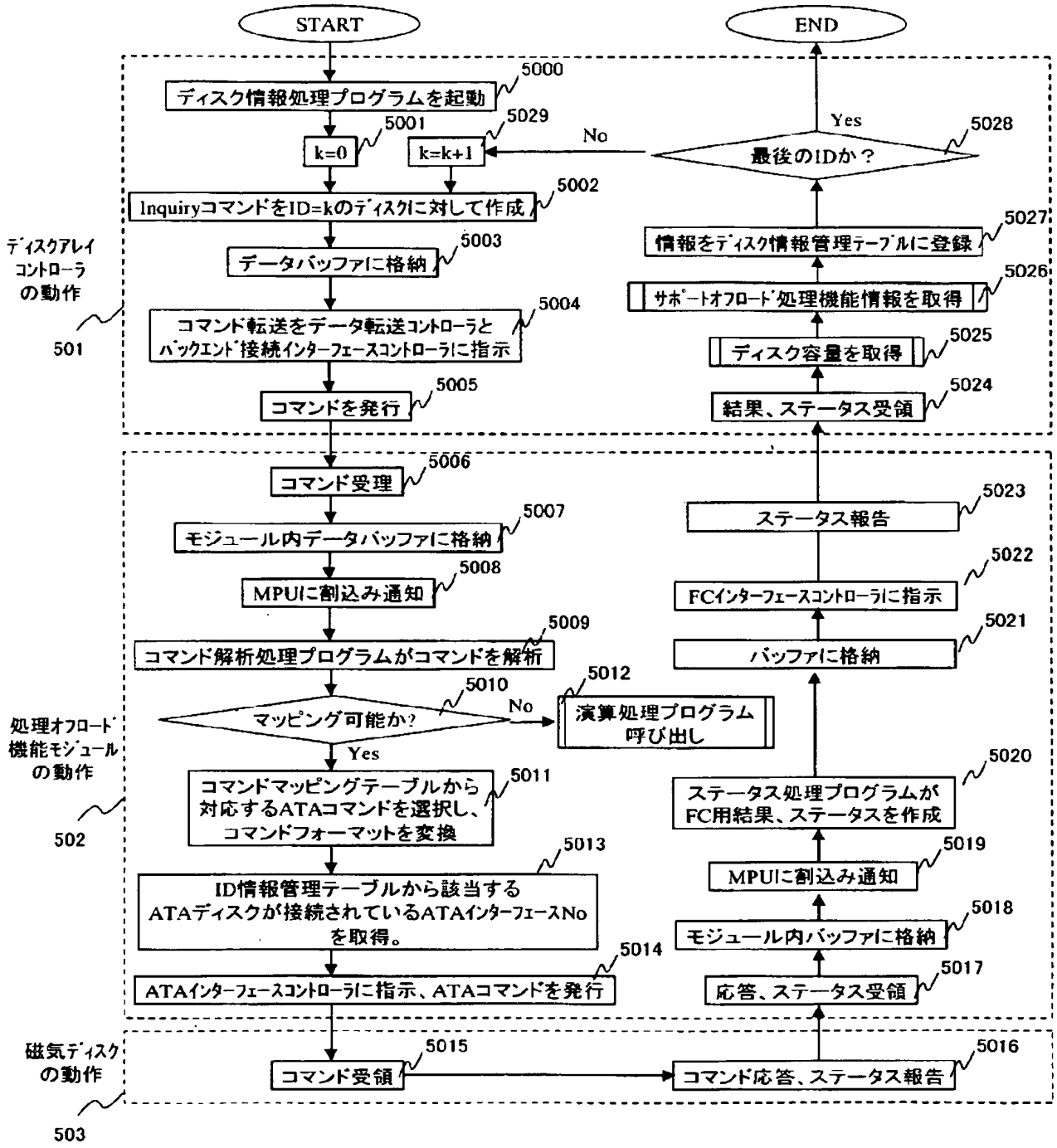
2063

ディスク情報管理テーブル

No (401)	ベンダ名 (402)	モデル名 (403)	シリアルNo (404)	容量 (405)	状態 (406)	オフロード機能 (407)			
						パリティ演算付ライト	RAIDフォーマット	オンラインベリファイ	ディスクコピー
#0	Hitachi	xxx-xxxxxxx	7F0450370A	200GB	良好	○	○	○	○
#1	Hitachi	xxx-xxxxxxx	7F0538432A	200GB	良好	○	○	○	○
#2	Hitachi	xxx-xxxxxxx	7F327532EB	200GB	良好	○	○	○	○
#3	Hitachi	xxx-xxxxxxx	8F3489601A	200GB	良好	○	○	○	○
#4	Hitachi	xxx-xxxyyyy	3W6014601D	200GB	良好	○	○	○	○
#5	Hitachi	xxx-xxxxxxx	2A3260112E	200GB	良好	○	○	○	○
#6	Hitachi	xxx-xxxxxxx	8E5436292F	200GB	良好	○	○	○	○
#7	Hitachi	xxx xxxxxxx	3A5236017F	200GB	良好	○	○	○	○
#27	Hitachi	xxx-xxxyyyy	3H2160143D	180GB	良好	○	×	○	○
#28	Hitachi	xxx-xxxyyyy	3B3541128A	180GB	良好	○	×	○	○

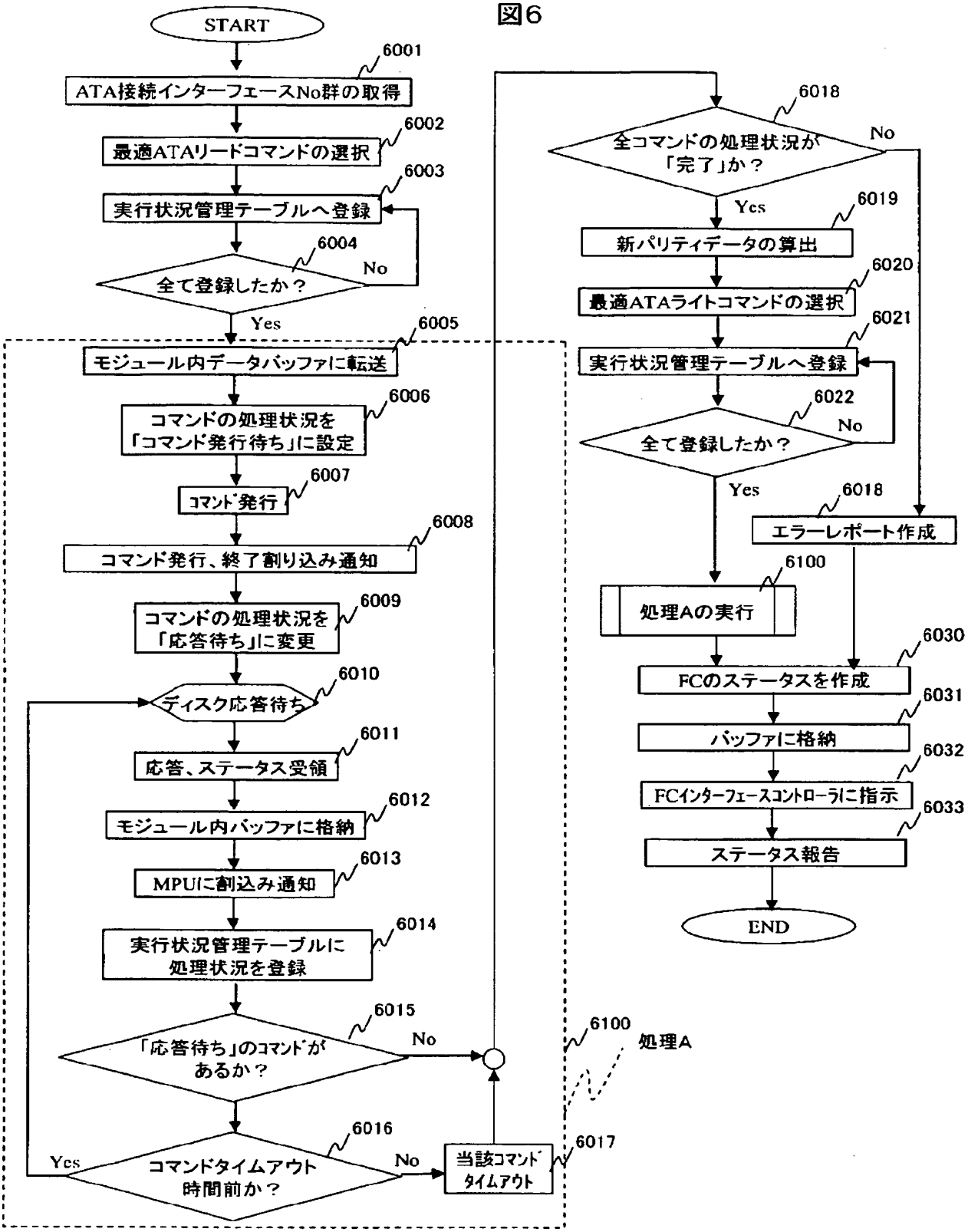
【図５】

図5


START
END
ディスク情報処理プログラムを起動
5000
k=0
5001
k=k+1
5029
No
最後のIDか？
5028
Yes
Inquiryコマンドを ID=kのディスクに対して作成
5002
データバッファに格納
5003
コマンド転送をデータ転送コントローラと バックエンド接続インターフェースコントローラに指示
5004
コマンドを発行
5005
情報をディスク情報管理テーブルに登録
5027
サポートオフロード処理機能情報を取得
5026
ディスク容量を取得
5025
結果、ステータス受領
5024
ディスクアレイ コントローラ の動作
501
コマンド受理
5006
モジュール内データバッファに格納
5007
MPUに割込み通知
5008
コマンド解析処理プログラムがコマンドを解析
5009
マッピング可能か?
5010
No
演算処理プログラム 呼び出し
5012
Yes
コマンドマッピングテーブルから 対応するATAコマンドを選択し、 コマンドフォーマットを変換
5011
ID情報管理テーブルから該当する ATAディスクが接続されているATAインターフェースNo を取得。
5013
ATAインターフェースコントローラに指示、ATAコマンドを発行
5014
処理オフロード 機能モジュール の動作
502
ステータス報告
5023
FCインターフェースコントローラに指示
5022
バッファに格納
5021
ステータス処理プログラムが FC用結果、ステータスを作成
5020
MPUに割込み通知
5019
モジュール内バッファに格納
5018
応答、ステータス受領
5017
コマンド受領
5015
コマンド応答、ステータス報告
5016
磁気ディスク の動作
503

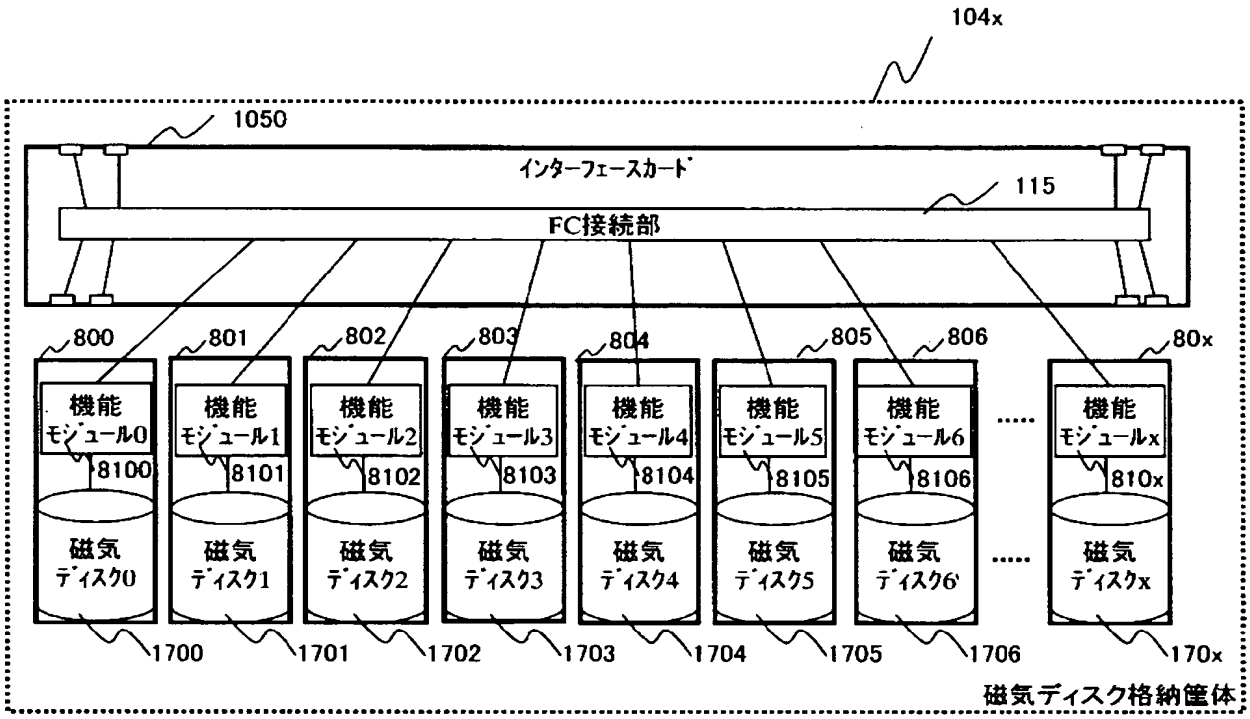
【図６】

図6


START
6001
ATA接続インターフェースNo群の取得
6002
最適ATAリードコマンドの選択
6003
実行状況管理テーブルへ登録
6004
全て登録したか？
No
Yes
6005
モジュール内データバッファに転送
6006
コマンドの処理状況を「コマンド発行待ち」に設定
6007
ｺﾏﾝﾄﾞ発行
6008
コマンド発行、終了割り込み通知
6009
コマンドの処理状況を「応答待ち」に変更
6010
ディスク応答待ち
6011
応答、ステータス受領
6012
モジュール内バッファに格納
6013
MPUに割込み通知
6014
実行状況管理テーブルに処理状況を登録
6015
「応答待ち」のｺﾏﾝﾄﾞがあるか？
No
6016
Yes
コマンドタイムアウト時間前か？
No
6017
当該ｺﾏﾝﾄﾞタイムアウト
6018
全コマンドの処理状況が「完了」か？
No
Yes
6019
新パリティデータの算出
6020
最適ATAライトコマンドの選択
6021
実行状況管理テーブルへ登録
6022
全て登録したか？
No
Yes
6018
エラーレポート作成
6100
処理Aの実行
6030
FCのステータスを作成
6031
バッファに格納
6032
FCインターフェースコントローラに指示
6033
ステータス報告
END
6100
処理A

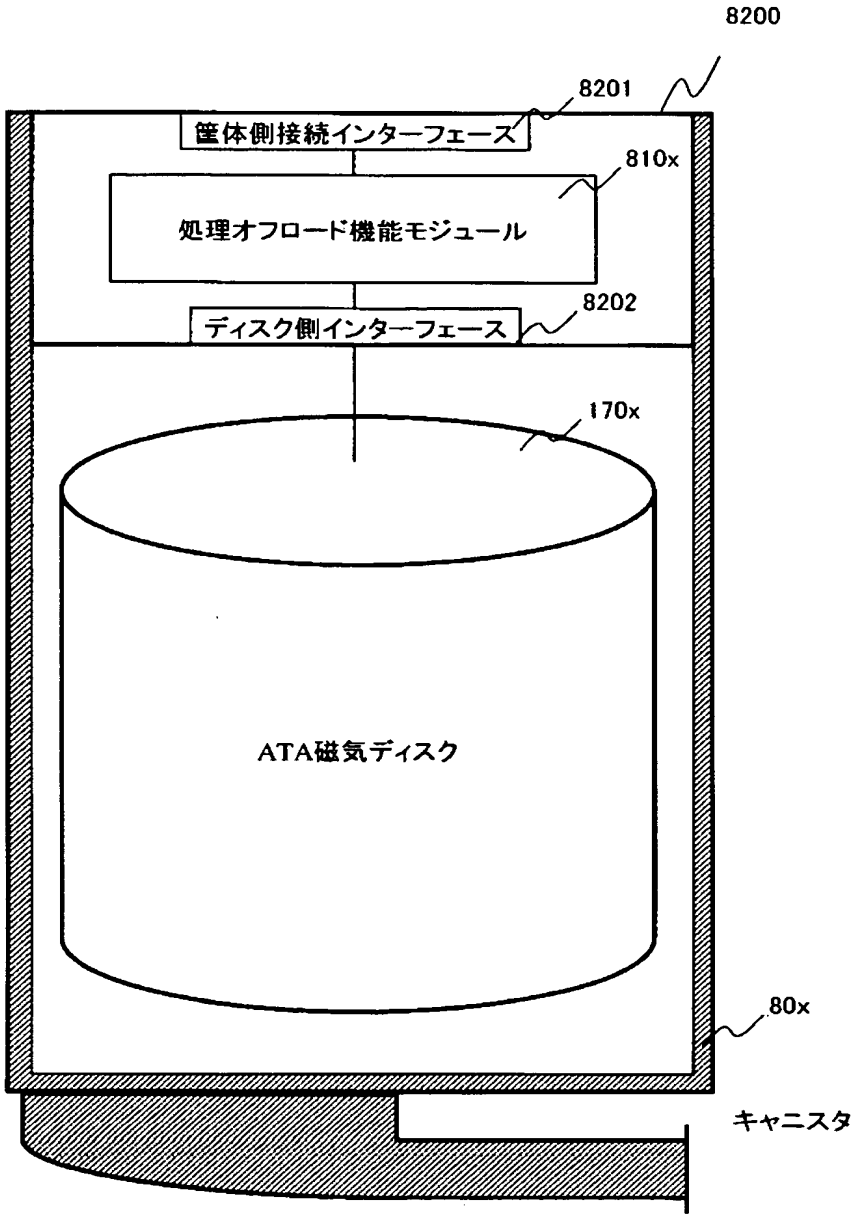
【図７】

図7


104x
1050
インターフェースカード
115
FC接続部
800
801
802
803
804
805
806
80x
機能モジュール0
機能モジュール1
機能モジュール2
機能モジュール3
機能モジュール4
機能モジュール5
機能モジュール6
機能モジュールx
8100
8101
8102
8103
8104
8105
8106
810x
磁気ディスク0
磁気ディスク1
磁気ディスク2
磁気ディスク3
磁気ディスク4
磁気ディスク5
磁気ディスク6
磁気ディスクx
1700
1701
1702
1703
1704
1705
1706
170x
磁気ディスク格納筐体

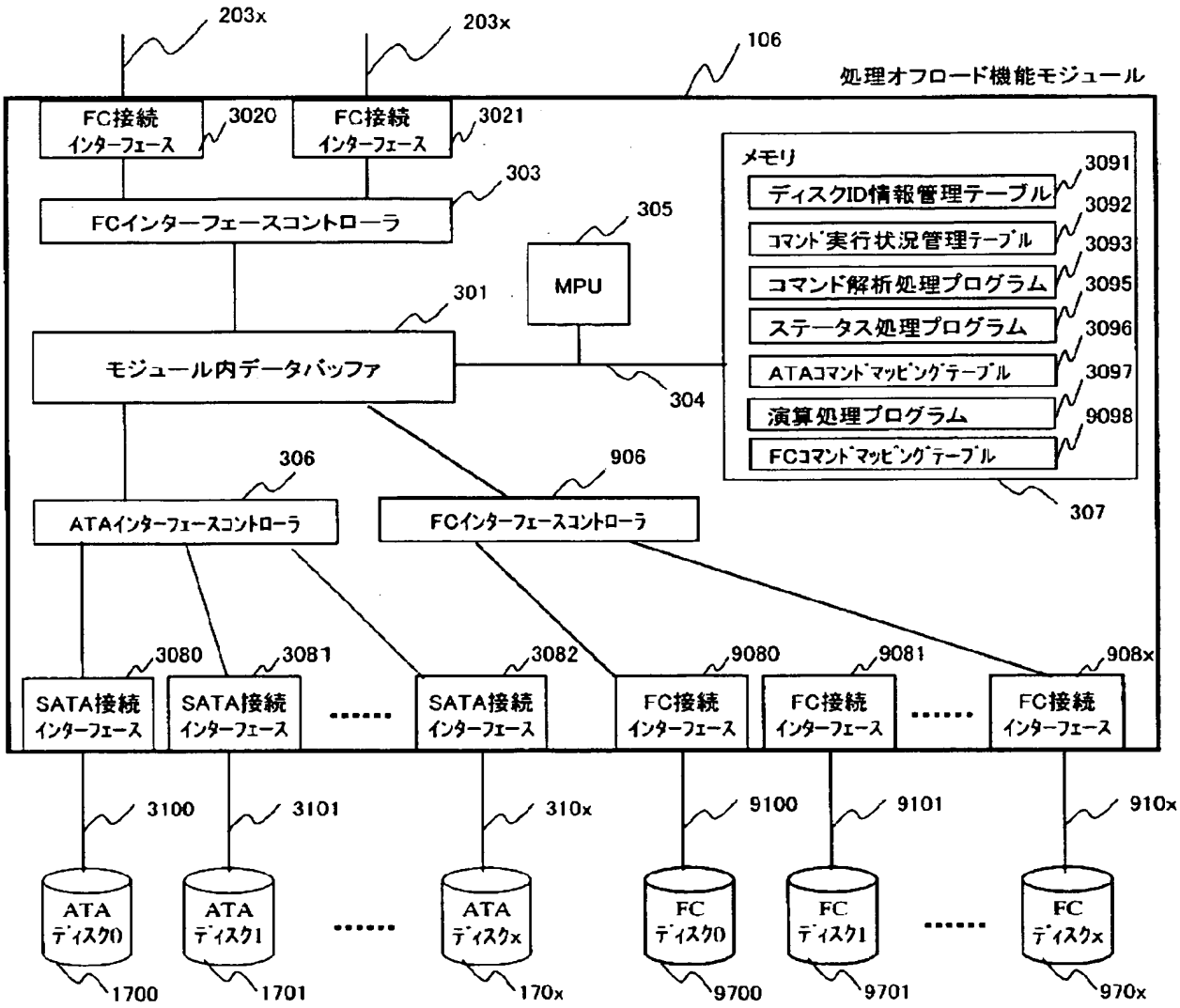
【図８】

図8


8200
8201
筐体側接続インターフェース
810x
処理オフロード機能モジュール
8202
ディスク側インターフェース
170x
ATA磁気ディスク
80x
キャニスタ

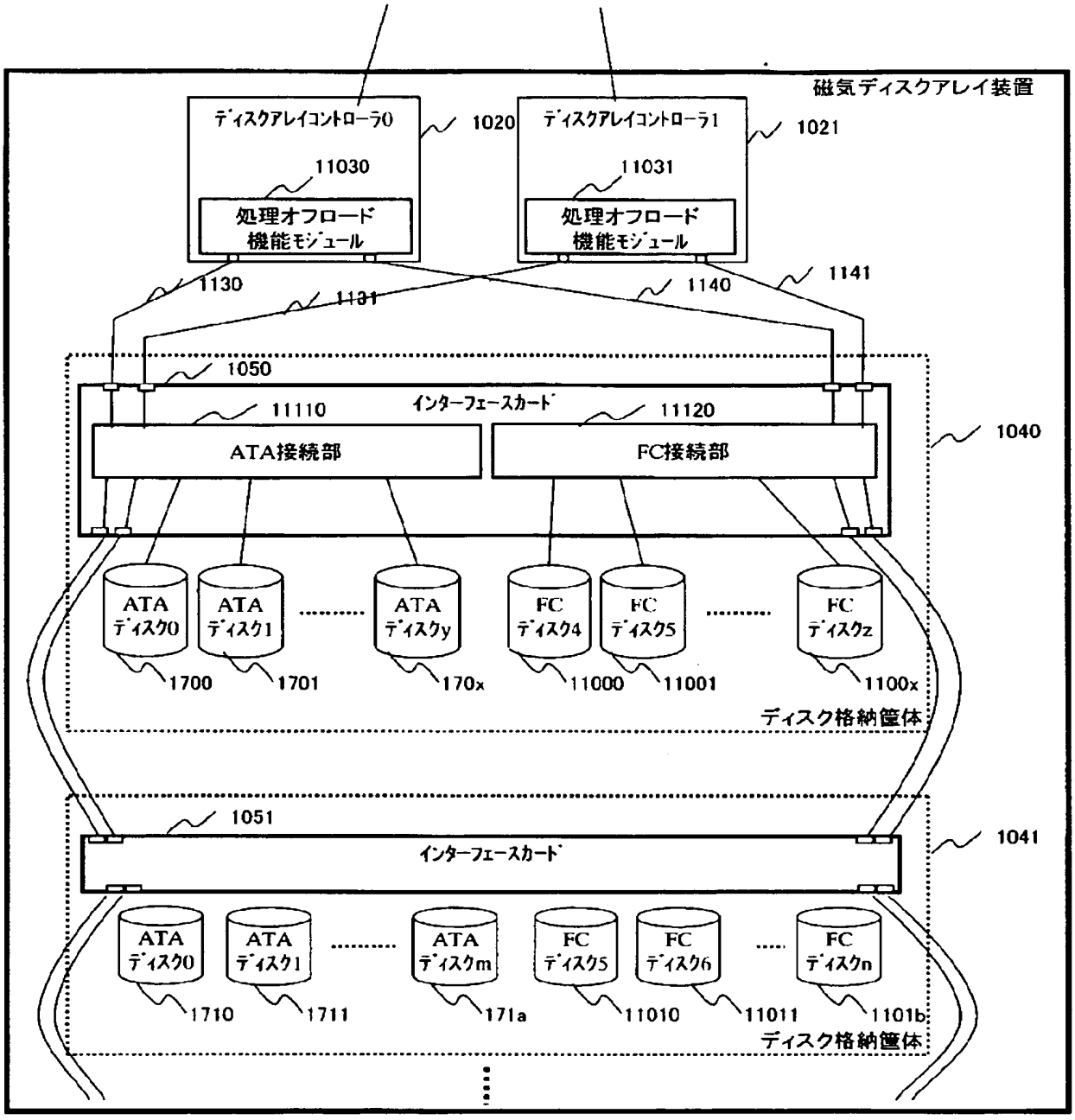
【図９】

図9


203x
203x
106
処理オフロード機能モジュール
FC接続 インターフェース
3020
FC接続 インターフェース
3021
303
FCインターフェースコントローラ
305
MPU
301
モジュール内データバッファ
304
メモリ
3091
ディスクID情報管理テーブル
3092
ｺﾏﾝﾄﾞ実行状況管理ﾃｰﾌﾞﾙ
3093
コマンド解析処理プログラム
3095
ステータス処理プログラム
3096
ATAｺﾏﾝﾄﾞﾏｯﾋﾟﾝｸﾞﾃｰﾌﾞﾙ
3097
演算処理プログラム
9098
FCｺﾏﾝﾄﾞﾏｯﾋﾟﾝｸﾞﾃｰﾌﾞﾙ
307
306
ATAインターフェースコントローラ
906
FCインターフェースコントローラ
3080
SATA接続 インターフェース
3081
SATA接続 インターフェース
3082
SATA接続 インターフェース
9080
FC接続 インターフェース
9081
FC接続 インターフェース
908x
FC接続 インターフェース
3100
3101
310x
9100
9101
910x
ATA ﾃﾞｨｽｸ0
ATA ﾃﾞｨｽｸ1
ATA ﾃﾞｨｽｸx
FC ﾃﾞｨｽｸ0
FC ﾃﾞｨｽｸ1
FC ﾃﾞｨｽｸx
1700
1701
170x
9700
9701
970x

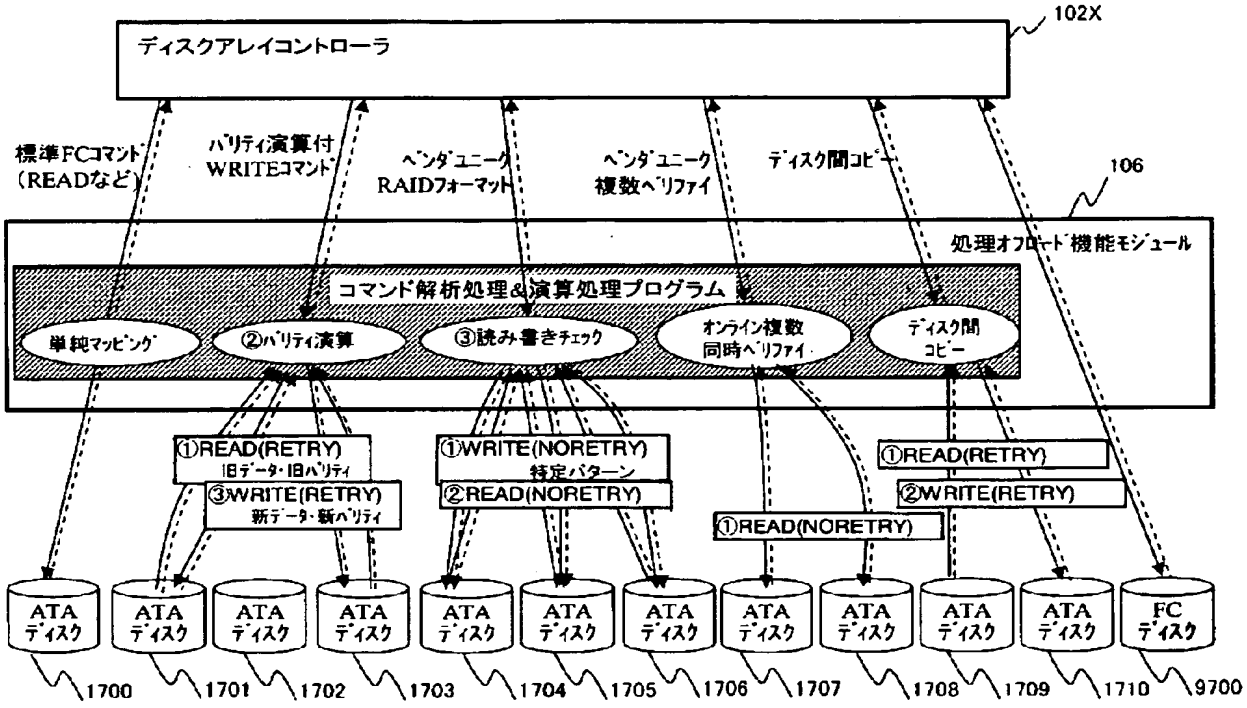
【図１０】

図10


磁気ディスクアレイ装置
ディスクアレイコントローラ0
1020
ディスクアレイコントローラ1
1021
11030
処理オフロード
機能モジュール
11031
処理オフロード
機能モジュール
1130
1131
1140
1141
1050
11110
インターフェースカード
11120
ATA接続部
FC接続部
1040
ATA
ディスク0
ATA
ディスク1
ATA
ディスクy
FC
ディスク4
FC
ディスク5
FC
ディスクz
1700
1701
170x
11000
11001
1100x
ディスク格納筐体
1051
インターフェースカード
1041
ATA
ディスク0
ATA
ディスク1
ATA
ディスクm
FC
ディスク5
FC
ディスク6
FC
ディスクn
1710
1711
171a
11010
11011
1101b
ディスク格納筐体

【図１１】

図11


102X
ディスクアレイコントローラ
標準FCコマンド
(READなど)
パリティ演算付
WRITEコマンド
ベンダユニーク
RAIDフォーマット
ベンダユニーク
複数ベリファイ
ディスク間コピー
106
処理オフロード機能モジュール
コマンド解析処理&演算処理プログラム
単純マッピング
②パリティ演算
③読み書きチェック
オンライン複数
同時ベリファイ
ディスク間
コピー
①READ(RETRY)
旧データ・旧パリティ
③WRITE(RETRY)
新データ・新パリティ
①WRITE(NORETRY)
特定パターン
②READ(NORETRY)
①READ(NORETRY)
①READ(RETRY)
②WRITE(RETRY)
ATA
ディスク
FC
ディスク
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
9700

【書類名】要約書
【要約】
【課題】ＡＴＡディスクの利用を可能としつつ、ＦＣ－ＡＴＡコマンド変換によるオーバーヘッドを削減し、最適なＡＴＡコマンドを生成し、ＩＤ等のリソースを無消費にして、オフロード処理を実現すること。
【解決手段】ＡＴＡ磁気ディスクと、ＡＴＡ磁気ディスクを制御するディスクアレイコントローラと、ディスクアレイコントローラとＡＴＡ磁気ディスクとの間の処理オフロード機能モジュールと、を備えた磁気ディスクアレイ装置であって、ディスクコントローラは、リード又はライト等の処理を行う標準処理ＦＣコマンドと、ベンダユニークなオフロード処理を行うオフロード処理ＦＣコマンドと、を出力し、処理オフロード機能モジュールは、標準処理ＦＣコマンドに対しコマンドマッピングして対応するＡＴＡコマンドをＡＴＡ磁気ディスクに発行し、オフロード処理ＦＣコマンドに対しＡＴＡプロトコルで最適処理となるＡＴＡコマンド群を用意する。
【選択図】図１

特願２００３－３９３９１２

出　願　人　履　歴　情　報

識別番号　　［０００００５１０８］

１．変更年月日　　１９９０年　８月３１日
［変更理由］　　新規登録
住　所　　東京都千代田区神田駿河台４丁目６番地
氏　名　　株式会社日立製作所